



# Conceptual engineering for truth: alethic properties and new alethic concepts

Kevin Scharp<sup>1</sup>

Received: 7 December 2016 / Accepted: 22 November 2019  
© The Author(s) 2020

## Abstract

What is the property of being true like? To answer this question, begin with a Canberra-plan analysis of the concept of truth. That is, assemble the platitudes for the concept of truth, and then investigate which property might satisfy them. This project is aided by Friedman and Sheard's groundbreaking analysis of twelve *logical* platitudes for truth. It turns out that, because of the paradoxes like the liar, the platitudes for the concept of truth are inconsistent. Moreover, there are so many distinct paradoxes that only small subsets of platitudes for truth are consistent. The result is that there is no property of being true. The failure of the Canberra plan analysis of the concept of truth, points the way toward a new methodology: a conceptual engineering project for the concept of truth. Conceptual engineering is assessing the quality of our concepts, and when they are found defective, offering new and better concepts to replace them for certain purposes. Still, there are many *alethic properties*, which are properties satisfied by reasonably large subsets of platitudes for the concept of truth. We can treat these alethic properties as a guide to the multitude of new alethic concepts, which are concepts similar to, but distinct from, the concept of truth. Any new alethic concept or team of concepts might be called on to replace the concept of truth. In particular, the concepts of ascending truth and descending truth are recommended, but the most important point is that we need a full-scale investigation into the space of alethic properties and new alethic concepts—that is, we need an Alethic Principles Project (APP).

**Keywords** Truth · Liar paradox · Conceptual engineering · Canberra plan · Methodology · Axiomatic theories of truth · Deflationism · Alethic principles project

---

✉ Kevin Scharp  
ks70@st-andrews.ac.uk

<sup>1</sup> Arché Philosophical Research Centre, University of St Andrews, St Andrews, UK

Over the last decade, there has been a fruitful interaction between those working on the nature of truth and those working on the paradoxes that affect truth, like the liar paradox. The liar paradox is that, by reflecting on sentences like ‘this very sentence is not true’, it is easy and quick to derive a contradiction: the sentence is both true and not true.<sup>1</sup> Although insulated from one another for decades, each tradition, *the nature of truth* and *the paradoxes*, has been invigorated by new unifying (or crossover) work contributed by many theorists.<sup>2</sup> A unifying project that draws from each tradition is our focus here.

In particular, there is a school of thought on the liar paradox that diagnoses it as a defect in our very concept of truth. We can call this the *inconsistency approach*. On the other hand, there has been a tremendous amount of work within the “nature of truth” tradition on the *metaphysics of truth*. Our question is: what might an *inconsistency theorist* about the paradoxes say about the *metaphysics of truth*?

A bit of background on each of these topics is in order, but first a word of caution. According to the usage here, *the property of being true* is the property designated by the English word ‘true’, which expresses *the concept of truth*. When something is true, it has the property of being true, whether it is a sentence, story, song, proposition, theory, utterance, prediction, or whatever. Likewise, anything that has the property of being true is true. The property (being true), the word (‘true’), and the concept (truth) are not to be confused.<sup>3</sup> Inconsistency theorists focus on the concept of truth and the word ‘true’, while the metaphysics of truth is about the property of being true. We shall question whether there is a property of being true, but *not* whether there is a word, ‘true’, or whether there is a concept of truth.

Inconsistency theorists claim that the principles essential to the concept of truth permit a competent reasoner to derive a contradiction (e.g., liar sentences are both true and not true). The inconsistency approach goes back to Alfred Tarski’s pioneering work in the 1930s, but it has really taken off since 2002, when Matti Eklund published “On Inconsistent Languages.”<sup>4</sup> Since then, much has been done to explore various options and provide details for what were once just suggestions. Because my own work falls in this tradition, I am especially interested in these exploratory endeavors. Much has been written about how concepts can be inconsistent, how these can be possessed and used, and what sorts of logics and semantics go best with an inconsistency approach. However, there is little on the property of being true from this perspective.

On the other hand, the central issue in the metaphysics of truth discussion is whether the property of being true is deflationary or substantive. Is being true more like a logical property (e.g., being necessary) or more like a scientific property (e.g.,

<sup>1</sup> See Künne (2003) for an overview of the *nature of truth tradition*, and see Field (2008) for an overview of the *paradox tradition*.

<sup>2</sup> See Beall and Armour-Garb (2005), Beall and Glanzberg (2008), and Scharp (2013) for overviews of unifying work between the two traditions.

<sup>3</sup> See Bar-On and Simmons (2007) and Asay (2013: ch. 1) for good examples of clarity on this.

<sup>4</sup> Tarski (1933), Eklund (2002); see also Burgess (2006) and Scharp (2013).

being a mammal)? Is the property of being true cut out to explain anything?<sup>5</sup> Almost all the work on the metaphysics of truth assumes without question that the concept of truth is just fine and there is a unique property of being true.

Our unifying topic is: *What does the metaphysical landscape look like for the inconsistency theorist?* This particular bridge over the two broad traditions has yet to be investigated in much detail. Is there a property of being true? If so, what is it like? Might there be more than one equally good candidate for this property? If so, which ones? How exactly can there be a property of being true if the concept of truth is defective? And, most importantly, if there is no such property, then are there any properties that are somewhat like what we thought the property of truth would be like? For this last question, think about it like this. We have lots of beliefs about what the property of being true should be like.<sup>6</sup> If there are no properties that satisfy these beliefs, then there still might be properties that come close enough to count as the property of being true. And even if there are no properties that come close enough, there will no doubt be properties that are somewhat similar in various ways to what we thought the property of truth would have to be like. What are these properties like?

In order to have a term for talking about all these properties at once without prejudging whether there is a unique property of being true, we can use ‘alethic property’. (‘Alethic’ just is an adjective synonymous with ‘pertaining to truth’.) Hence, an *alethic property* is any property that is similar to what we think the property of truth would have to be like. With this terminology in hand, we can say: we are investigating the class of alethic properties, and in particular we are investigating the alethic properties from the point of view of an inconsistency approach.

Our investigation is significant even to those who reject the inconsistency approach because we shall begin *without* prejudging the matter of whether the concept of truth itself is defective. We shall begin by assuming a neutral framework that is ubiquitous in contemporary analytic philosophy: the Canberra plan, which is just a useful term for the philosophical methodology according to which one first collects the relevant platitudes for some concept and then one searches for a unique thing in the world that satisfies them.<sup>7</sup> By following this strategy, we arrive at a novel and powerful argument for the inconsistency view, which appeals to a much wider range of considerations than anything in the literature now.

This strategy for investigating the alethic properties has consequences for the literature on the metaphysics of truth as well. It has the virtue of providing an illuminating framework for investigating the alethic properties. Moreover, our strategy can be encapsulated in the directive: *first* determine which alethic properties there

<sup>5</sup> See Beall and Armour-Garb (2005) for a helpful summary and classic papers on deflationism. See Wyatt (2015) for an overview of the discussion about whether the property of being true is deflationary.

<sup>6</sup> I am not going to be precise about this matter yet. I use phrases like ‘beliefs about what truth should be like’ and ‘beliefs about what truth must be like’ and ‘beliefs about what is essential to truth’ interchangeably. See Section One for further discussion of this topic and the nature of the platitudes for truth.

<sup>7</sup> The Canberra plan owes much to Lewis (1970) and Jackson (1997).

are, *then* figure out whether they are deflationary or substantive or otherwise. This directive is the opposite of how almost everyone in this literature proceeds.

I shall argue that *there is no property of being true* because no property comes close enough to satisfying the platitudes for truth. Even if there is no property of being true, there is still a worthy philosophical project of identifying the properties that are similar to what we thought being true was like. Why is it worthy? The alethic properties are the ones that come closest to satisfying the beliefs we have about what the property of being true should be like. If we want to represent the world accurately, then it is *these* properties we should be trying to represent with our talk and thought. Indeed, it seems reasonable if we have certain purposes to require that the concepts we use really denote something out there in the world. That is, for certain purposes, we might want to avoid concepts that are so defective that there is nothing in the world for them pick out. Imagine that a person, Deandra, has some goal that requires that she avoid using defective concepts (or at least ones that are so defective they do not refer to anything). Then Deandra should not use the concept of truth for that purpose. Which concept or concepts should she use in place of the concept of truth? The answer to this question turns on which of the alethic properties behaves most similarly to *what we thought was* the behavior of the property of being true. In simpler terms: which alethic properties will do what we thought the property of being true could do? As of right now, this is an open question.

Identifying how we ought to change our conceptual scheme is one topic in the rapidly growing new field of *conceptual engineering*.<sup>8</sup> Once we see that there is no property of being true and we start looking around for which concepts to use in place of our defective concept of truth, we are doing conceptual engineering. It turns out that, right now, we do not have enough information about which alethic properties exist to complete a satisfactory conceptual engineering project with respect to truth. So we cannot answer the question: which non-defective concepts should we use instead of the concept of truth? We do not currently know enough to answer it. In response to this result, I suggest that we, the community of theorists with an interest in the paradoxes affecting truth, ought to investigate which alethic properties exist by identifying which sets of platitudes for truth are consistent. We can call this the *Alethic Principles Project* (APP), and the paper ends by trying to motivate you to take it up because we need your help.

## 1 The plan

We start by following the Canberra plan, certainly one of the most popular philosophical methodologies at the present time. This strategy will eventually be abandoned because there turns out to be no property even close to what we think of as the property of being true. We then switch to conceptual engineering to investigate

<sup>8</sup> For overviews of conceptual engineering, see Burgess and Plunkett (2013a, b), Cappelen (2018), and Scharp (2020).

which new concepts like truth we might adopt, given which aletheic properties turn out to exist.<sup>9</sup>

The Canberra plan is one way of pursuing conceptual analysis—the methodology for which analytic philosophy is named. Instead of seeking a single definition for some philosophical term, the Canberra plan begins with platitudes about the topic in question.<sup>10</sup> Roughly, a platitude is a claim that seems obviously true, self-evident, or commonsensical. Some Canberra planners are more precise about platitudes, while others leave it at an intuitive level. The focus on platitudes gives the Canberra plan much more flexibility than traditional analysis since the platitudes need not constitute anything like a definition or even necessary or sufficient conditions.

The platitudes for a given concept might be analytic (i.e., true in virtue of their meaning alone), but one can follow the Canberra plan even if one takes platitudes to be simply uncontroversial principles or bits of common sense. One arrives at different versions of the Canberra plan by imposing stricter or looser standards on what counts as a platitude. For our purposes, we do not need to specify much of a standard because all the platitudes we care about are intuitively uncontroversial. We can say that all the platitudes we consider would be accepted by the vast majority of competent users of ‘true’ as obvious. Moreover, they all count as *constitutive* of the concept of truth in a certain sense: if a person rejects one of these principles in a conversation, that is a pro tanto reason to think that the person’s word ‘true’ does not express the concept of truth. I have developed this notion of constitutive principles elsewhere, but it will not play a role in what follows.<sup>11</sup>

After assembling the platitudes, we then engage in metaphysics by finding something in the world that does the best job of satisfying the platitudes. If the world cooperates, then the platitudes will be satisfied by a unique thing in the world. However, even if nothing perfectly fits the platitudes, something might fit them relatively well—well enough to say that it is what the term in question is ultimately about. In either of these cases, one might continue the investigation by considering whether that thing that fits the platitudes is fundamental or derivative, and if it is derivative, how it relates to the fundamental level of reality. These are the cases in which the Canberra plan is most fruitful and philosophically rich. In other cases, it might be that multiple things do an equally good job of satisfying the platitudes or perhaps nothing even comes close to satisfying them. Canberra planners spend little time considering these outcomes, and it is not clear what to do in these cases. Unfortunately, we will have to think more about these cases because truth turns out to be in this category.

The Canberra plan is a perfect starting place for our inquiry into what an inconsistency theorist ought to say about the metaphysics of truth. In fact, the Canberra

<sup>9</sup> See (Braddon-Mitchell and Robert 2009) and Johnston and Leslie (2012) for discussions of the Canberra Plan.

<sup>10</sup> There are other ways of pursuing analytic projects besides the Canberra plan, but we have good reason to think they would deliver the same results. There are other methodologies as well, but they are beyond the scope of the paper.

<sup>11</sup> See Scharp (2013, 2018).

plan seems to encapsulate the common sense idea that we treat the platitudes for *the concept of truth* as a guide to *the property of being true*. That is, the platitudes that many or most of us treat as constitutive of the concept of truth are assumed to be principles governing the property of being true as well. For example, we think that if snow is white, then ‘snow is white’ is true. When thinking about the property of being true, we often unreflectively treat this as a principle of the property of being true: if snow is white, then ‘snow is white’ has the property of being true.

## 2 Truth platitudes

What are the platitudes for truth? There are so many to choose from, but we can begin with Michael Lynch’s list, which is representative.

(Objectivity)	The belief that p is true if, and only if, with respect to the belief that p, things are as they are believed to be.
(BS)	The belief that p is true if and only if p.
(TS)	The proposition that p is true if, and only if, p.
(Warrant Independence)	Some beliefs can be true but not warranted and some can be warranted without being true
(Norm of Belief)	It is <i>prima facie</i> correct to believe that p if and only if the proposition that p is true.
(End of Inquiry)	Other things being equal, true beliefs are a worthy goal of inquiry [Lynch (2009: ch. 1)].

Lynch is right that each of these principles has a good case for being called a platitude for truth. Each one seems central to how we think about truth. If we begin with these as our platitudes, then we can move on to investigate which property might uniquely satisfy them. That is, we can look around for which property is *the* property of being true.

However, the problem with using the Lynch platitudes is that they are not as precise as we might like. This is not a criticism of Lynch—he recognizes that many of the central platitudes concerning truth are somewhat fuzzy or vague principles linking truth to other concepts like belief, warrant, normativity, and inquiry. But there are many more platitudes concerning truth that *are* precise, and few people bother to even enumerate them because they seem so innocuous. I am thinking in particular of the *logical platitudes*, which relate truth to various logical expressions. By focusing on them, we can utilize the tools of logic to investigate the extent to which the logical platitudes form a consistent set. Further, if the logical platitudes fail to pick out a unique best option for the property of being true, then the fuzzier platitudes Lynch enumerates will fail as well. The reason is that most or all of the Lynch platitudes presuppose the truth of various logical platitudes. For example, if the Objectivity platitude is true, then, as Lynch observes, one can derive instances of Schema T from it (e.g., ‘ $2 + 2 = 4$ ’ is true iff  $2 + 2 = 4$ ).

Moreover, and *this is the most important point*, using the Lynch platitudes in a Canberra-plan analysis of truth turns out to obscure the fine-grained difficulties

associated with identifying which property does the best job of satisfying the platitudes. In other words, we will see that the platitudes for truth are inconsistent, as evidenced by the liar paradox, but there are many *many* more paradoxes associated with truth, and each one is the result of a distinct subset of the platitudes. It turns out that the Lynch platitudes do not allow us to distinguish enough of these paradoxes because many of these paradoxes involve platitudes that are not explicitly among the Lynch platitudes. As a result, if we choose the Lynch platitudes, then we are led to think that the concept of truth is in far better shape than it actually is. Instead, we need to look closely at the *logical platitudes about truth* so we can get a better picture of the defects in our concept of truth.

### 3 Logical platitudes about truth

What exactly are the logical platitudes about truth? They include all and only the platitudes that contain only occurrences of a truth predicate and logical expressions like negation, disjunction, and universal quantification. The vast majority of these logical platitudes about truth are studied with the help of formal logical systems that are rather technical. Many of the technical details will not matter for our purposes here. Again, we are not engaged in a logical or technical project. Instead, we are interested in establishing a particular connection between technical work in logic on the paradoxes affecting truth and more mainstream work on the nature of truth. In particular, we are using some technical logical results to argue that the concept of truth is *so* defective that it does not pick out any property whatsoever. What follows is a presentation of some of this technical material for a more general audience. I hope that theorists familiar with the complexities of the technical material are patient with my attempts to summarize it for a non-technical audience. This sort of accessible summary inevitably involves some vagueness, distortion and sloppiness. My aim is to ignore nothing that makes a difference to the project with which I am engaged.

One of the most famous logical platitudes about truth (just ‘logical platitudes’ from here on) is:

- (Schema T)  $\langle p \rangle$  is true if and only if  $p$ . [See Tarski (1933). There are many ways of formulating this principle and others like it. For example, it is related to:  
(Intersubstitutability) Occurrences of  $p$  and ‘ $p$  is true’ can substituted in any sentence  $q$  without changing the truth value of  $q$ .]

In this schema, ‘ $p$ ’ is a sentential variable, which means the instances of (Schema T) have a sentence in place of ‘ $p$ ’, and the angle brackets are a naming device associated with sentential variables. That is, in the instances of the (Schema T), a name or description of the sentence that goes in for ‘ $p$ ’ goes in for ‘ $\langle p \rangle$ ’. Again, (Schema T) is inconsistent when we allow certain sentences in for ‘ $p$ ’ (e.g., liar sentences)

and we work in a setting of classical logic or any of the most familiar alternatives to classical logic like intuitionistic logic or the logic of relevant implication.<sup>12</sup>

For our purposes, the difference between schemata in the metalanguage and quantified sentences of the object language will not matter.<sup>13</sup> We are interested primarily in *natural language*, its truth predicate, the concept of truth this predicate expresses, and the paradoxes to which this concept gives rise. In what follows, I present the logical platitudes for truth as schemata only because these are easier for non-experts to read and understand. Our project is not to prove some new technical result, which would require extreme care with respect to these matters, but rather *to show the philosophical significance* (e.g., there is no property of being true) of some technical results.

This is *usually* the point at which those of us who are inconsistency theorists about truth conclude our case: (Schema T) is both central to the concept of truth *and* inconsistent in reasonable logics. Hence, the concept of truth is defective.<sup>14</sup> That is, the typical inconsistency theorist puts all the weight on a *single* logical platitude, (Schema T), and a *single* paradox, the version of the liar that shows (Schema T) is inconsistent. I want to be clear that I agree with this conclusion, but this standard argument for the inconsistency of truth, by itself, is not enough. It does not give us anything like a comprehensive picture of the defect in our concept of truth. *Inconsistency theorists have, so far, only seen the tip of the iceberg.*

We need to consider a wide range of logical platitudes for truth and investigate all the paradoxes that can be found among them. Only once we have this broad picture of the various defects in our concept of truth can we begin a genuine investigation into which subsets of alethic platitudes are consistent, and thereby focus on which alethic properties exist.

We can start our exploration by separating each direction of the biconditional in (Schema T) and then consider several weakenings of these conditional formulations as in the following *disquotational platitudes*<sup>15</sup>:

<sup>12</sup> These logics are motivated by considerations about logic; e.g., the paradoxes of implication for relevant logic and warrant-transfer for intuitionistic logic. There are logics in which (Schema T) is not trivial, like BX. But these logics are motivated not by considerations about logic, but by how to avoid the paradoxes of truth. Of course, insofar as BX is a relevant logic, it is motivated by avoiding the paradoxes of implication. However, the vast majority of proponents of relevant logics advocate stronger relevant logics like R or E, and these logics play no role in approaches to the alethic paradoxes because they are not weak enough to formulate plausible theories of truth non-trivially.

<sup>13</sup> There are myriad technical issues in the neighborhood that matter for logical investigations of formal languages, but do not matter for our purposes. Examples include the difference between individual and propositional quantification, which is a matter of the kinds of variables that are in the object language, and the difference between objectual and substitutional quantification, which is a matter of how the quantifiers and variables are interpreted. See Kripke (1976), Hugly and Sayward (2002), Hand (2007), and Georgi (2015). Moreover, there is the issue of how to explain quantification into quotation devices. See Belnap and Grover (1973) and Halbach (2011: pp. 29–36). See Field (2004), Halbach and Horsten (2008), Horsten (2011), and Cieśliński (2017) on the contrast between (Schema T) and other disquotational principles.

<sup>14</sup> Eklund (2002), Burgess (2006), and my own presentation in Scharp (2013).

<sup>15</sup> Each logical platitude for truth is given as a schematic formula and as an English principle. The English translations of these formulas can be distorting in various ways, but these do not matter for my purposes.



(T-In)	If $p$ , then $\langle p \rangle$ is true $p \rightarrow T\langle p \rangle$
(T-Out)	If $\langle p \rangle$ is true, then $p$ $T\langle p \rangle \rightarrow p$
(T-Intro)	From $p$ infer $\langle p \rangle$ is true $p \vdash T\langle p \rangle$
(T-Elim)	From $\langle p \rangle$ is true infer $p$ $T\langle p \rangle \vdash p$
( $\neg$ T-Intro)	From $p$ 's negation infer $\langle p \rangle$ is not true $\neg p \vdash \neg T\langle p \rangle$
( $\neg$ T-Elim)	From $\langle p \rangle$ is not true infer $p$ 's negation $\neg T\langle p \rangle \vdash \neg p$
(T-Enter)	If it is provable that $p$ , then it is provable that $\langle p \rangle$ is true $\vdash p \rightarrow \vdash T\langle p \rangle$
(T-Exit)	If it is provable that $\langle p \rangle$ is true, then it is provable that $p$ $\vdash T\langle p \rangle \rightarrow \vdash p$
( $\neg$ T-Enter)	If it is provable that not $p$ then it is provable that $\langle p \rangle$ is not true $\vdash \neg p \rightarrow \vdash \neg T\langle p \rangle$
( $\neg$ T-Exit)	If it is provable that $\langle p \rangle$ is not true, then it is provable that not $p$ $\vdash \neg T\langle p \rangle \rightarrow \vdash \neg p$
(T-Rep)	If $\langle p \rangle$ is true, then $\langle \langle p \rangle \text{ is true} \rangle$ is true $T\langle p \rangle \rightarrow T\langle T\langle p \rangle \rangle$
(T-Del)	If $\langle \langle p \rangle \text{ is true} \rangle$ is true, then $\langle p \rangle$ is true $T\langle T\langle p \rangle \rangle \rightarrow T\langle p \rangle$

(T-In) and (T-Out) constitute each conditional direction of (Schema T). (T-Elim) and (T-Intro) are inference rules, not conditionals. (T-Intro) is weaker than (T-In) and (T-Elim) is weaker than (T-Out).<sup>16</sup>

Weaker still are the derivation rules (T-Enter) and (T-Exit). The former says that if one can derive a sentence, then one can derive the truth attribution to that sentence, and the latter says the converse. These are even weaker than the inference rules because inference rules can be used in hypothetical contexts (e.g., reductio arguments and conditional proof arguments), but the derivation rules cannot. The derivation rules can be utilized only for formulas proven by the theory in question (i.e., the theorems). One can think of the derivation rules as special cases of the inference rules restricted to the theorems of a theory. The derivation rules, (T-Enter) and (T-Exit) are weak enough to be consistent in classical logic. Finally, (T-Rep) and (T-Del) are like the two directions of (Schema T) but restricted to atomic truth attributions.

<sup>16</sup> In certain logics weaker than classical logic, (T-In) and (T-Out) are inconsistent, but (T-Intro) and (T-Elim) are consistent.; see Field (2008: ch 11) on weakly classical approaches. However, (T-In) and (T-Intro) are equivalent in classical logic, as are (T-Out) and (T-Elim).

All the above principles come from reflecting on (Schema T). The next batch of platitudes involve the truth predicate distributing over logical vocabulary; they are *compositional platitudes*.

- ( $\neg$ -Imb) If a sentence is not true, then its negation is true  
 $\neg T\langle p \rangle \rightarrow T\langle \neg p \rangle$
- ( $\neg$ -Exc) If a sentence's negation is true, then the sentence is not true  
 $T\langle \neg p \rangle \rightarrow \neg T\langle p \rangle$
- ( $\wedge$ -Imb) If two sentences are true, then their conjunction is true  
 $T\langle p \rangle \wedge T\langle q \rangle \rightarrow T\langle p \wedge q \rangle$
- ( $\wedge$ -Exc) If a conjunction is true, then its conjuncts are true.  
 $T\langle p \wedge q \rangle \rightarrow T\langle p \rangle \wedge T\langle q \rangle$
- ( $\vee$ -Imb) If two sentences are true, then their disjunction is true  
 $T\langle p \rangle \vee T\langle q \rangle \rightarrow T\langle p \vee q \rangle$
- ( $\vee$ -Exc) If a disjunction is true, then one of the disjuncts is true.  
 $T\langle p \vee q \rangle \rightarrow T\langle p \rangle \vee T\langle q \rangle$
- ( $\rightarrow$ -Imb) If one truth claim implies another, then their conditional is true  
 $(T\langle p \rangle \rightarrow T\langle q \rangle) \rightarrow T\langle p \rightarrow q \rangle$
- ( $\rightarrow$ -Exc) If a conditional is true, then if the antecedent is true then the consequent is true  
 $T\langle p \rightarrow q \rangle \rightarrow (T\langle p \rangle \rightarrow T\langle q \rangle)$

For each connective, there are two principles, each relating the connective in question, the truth predicate, and sentences. Note that ( $\neg$ -Imb) and ( $\neg$ -Exc) have a special significance because ( $\neg$ -Imb) is logically equivalent to the claim that for every sentence, either it or its negation is true (sometimes called *truth completeness*), and ( $\neg$ -Exc) is logically equivalent to the claim that no sentence is both true and not true (sometimes called *truth consistency*). These are listed separately below.

- (T-Comp) Either a sentence or its negation is true  
 $T\langle p \rangle \vee T\langle \neg p \rangle$
- (T-Cons) It is not the case that a sentence and its negation are true.  
 $\neg (T\langle p \rangle \wedge T\langle \neg p \rangle)$

It might not be obvious that the negation principles have these consequences, but they play an important role in what follows.

The next set of platitudes concerns the relationships between logical relations among sentences and logical relations of sentences in the extension of the truth predicate; we can call them *implication platitudes*:

- (MPC) If a conjunction of sentences implies another sentence, then the conjunction of truth claims to the conjuncts implies the truth claim to the consequent.  
 $(p_1 \wedge \dots \wedge p_n \rightarrow q) \rightarrow (T\langle p_1 \rangle \wedge \dots \wedge T\langle p_n \rangle \rightarrow T\langle q \rangle)$
- (SPC) If one sentence implies another, then if the antecedent is true, then the consequent is true.  
 $(p \rightarrow q) \rightarrow (T\langle p \rangle \rightarrow T\langle q \rangle)$

- (Sub-In) If one sentence is equivalent to another, then their truth claims are equivalent.  
 $p \leftrightarrow q \rightarrow T\langle p \rangle \leftrightarrow T\langle q \rangle$
- (MPT) If a conjunction of truth claims to some sentences implies another truth claim, then the conjunction of those sentences as implies the other sentence.  
 $(T\langle p_1 \rangle \wedge \dots \wedge T\langle p_n \rangle) \rightarrow T\langle q \rangle \rightarrow (p_1 \wedge \dots \wedge p_n \rightarrow q)$
- (SPT) If one truth claim to a sentence implies another, then the first sentence implies the second.  
 $(T\langle p \rangle \rightarrow T\langle q \rangle) \rightarrow (p \rightarrow q)$
- (Sub-Out) If two truth claims to sentences are equivalent, then the sentences are equivalent.  
 $T\langle p \rangle \leftrightarrow T\langle q \rangle \rightarrow (p \leftrightarrow q)$

(MPC), which stands for MultiPremise Closure, entails (SPC), which entails (Sub-In); likewise, (MPT), which stands for MultiPremise Tracking, entails (SPT), which entails (Sub-Out). The closure platitudes say that the sentences in the extension of the truth predicate have the same implication structure as the sentences of the language, while the tracking platitudes say that the sentences of the language have the same implication structure as the sentences in the extension of the truth predicate.

Another set of platitudes we consider seem very intuitive because they say that logical tautologies are true, that logical contradictions are not true, and that the set of truths is closed under various logical inference rules; we can call them *inferential platitudes*.

- (Taut) Tautologies are true  
 $T\langle p \rangle$  for  $p$  a tautology
- (Contra) Contradictions are not true.  
 $\neg T\langle p \rangle$  for  $p$  a contradiction
- (MP) Truth is closed under modus ponens  
 $(T\langle p \rightarrow q \rangle \wedge T\langle p \rangle) \rightarrow T\langle q \rangle$
- (MT) Truth is closed under modus tollens  
 $(T\langle \neg q \rangle \wedge T\langle p \rightarrow q \rangle) \rightarrow T\langle \neg p \rangle$
- (DN) The truths are closed under double negation elimination  
 $T\langle \neg \neg p \rangle \rightarrow T\langle p \rangle$
- (HS) The truths are closed under hypothetical syllogism  
 $T\langle p \rightarrow q \rangle \wedge T\langle q \rightarrow r \rangle \rightarrow T\langle p \rightarrow r \rangle$
- (Red) The truths are closed under reductio  
 $(T\langle p \rangle \rightarrow T\perp) \rightarrow T\langle \neg p \rangle$
- (Dil) The truths are closed under dilemma  
 $T\langle p \rightarrow r \rangle \wedge T\langle q \rightarrow r \rangle \wedge T\langle p \vee q \rangle \rightarrow T\langle r \rangle$
- (CP) The truths are closed under conditional proof  
 If  $T\langle p \rangle \vdash T\langle q \rangle$  then  $\vdash T\langle p \rightarrow q \rangle$

It is rather intuitive to think that good deductive reasoning takes one from true claims to other true claims, that tautologies are true, and that contradictions are not true. Note also that (MP) is equivalent to ( $\rightarrow$ -Exc),  $T\langle p \rightarrow q \rangle \rightarrow (T\langle p \rangle \rightarrow T\langle q \rangle)$ .

Dialetheists claim that some contradictions are true, but they should be perfectly happy with (Contra) being a platitude.<sup>17</sup> For the most prominent dialetheists, true contradictions are the result of having inconsistent analytic principles. So as long as the set of platitudes for truth is inconsistent—and we are going to see that it is very, very hard to get a consistent set of platitudes that has more than a handful of members—the dialetheist can accept that most people take for granted ‘true contradiction’ is a contradiction in terms. Or, in other words, dialetheism is no reason to think that (Contra) is not a platitude.

The next group of platitudes concerns the interaction between truth and logical quantifiers. The *quantificational platitudes* are:

- (U-Imb) If all instances of a universal generalization are true, then the universal generalization is true.  
 $(\forall x)T\langle\phi(x)\rangle \rightarrow T\langle(\forall x)\phi(x)\rangle$
- (U-Exc) If a universal generalization is true, then all instances of the universal generalization are true.  
 $T\langle(\forall x)\phi(x)\rangle \rightarrow (\forall x)T\langle\phi(x)\rangle$
- (E-Imb) If some instance of an existential generalization is true, then the existential generalization is true.  
 $(\exists x)T\langle\phi(x)\rangle \rightarrow T\langle(\exists x)\phi(x)\rangle$
- (E-Exc) If an existential generalization is true, then some instance of the existential generalization is true.  
 $T\langle(\exists x)\phi(x)\rangle \rightarrow (\exists x)T\langle\phi(x)\rangle$

The formulas are more complicated in these cases because of the quantifiers. In these formulas, ‘ $\phi(x)$ ’ serves as a metalinguistic variable ranging over sentences of the language that might have ‘ $x$ ’ as a free variable. Because of the complexities surrounding quantifiers, the English glosses of these platitudes are a bit rough, but these details won’t matter for us.

Finally, we turn to logical platitudes concerned with the truth of other logical platitudes. Two stand out:

- (T-TIn) (T-In) is true.  
 $T\langle p \rightarrow T\langle p \rangle \rangle$
- (T-TOut) (T-Out) is true.  
 $T\langle T\langle p \rangle \rightarrow p \rangle$

One can formulate platitudes like these for each of the other logical platitudes above. It makes sense to focus on them because we want our theory of truth to be true. In particular, self-refuting theories (i.e., those that entail they are not true) are unacceptable. As we will see, (T-TOut) is special because it plays a central role in a paradox discussed in Section Five.

There are plenty of other logical platitudes for truth, but the ones listed so far seem to be the most central, and they are the ones that have received the most attention.

<sup>17</sup> See Priest (2006) and Beall (2009) for dialetheist approaches to truth and the paradoxes.

Our focus on logical platitudes for truth in our Canberra plan for truth can be interpreted as a strategy of letting *axiomatic theories of truth* be our guide to the alethic properties.<sup>18</sup> We have chosen to focus only on the logical platitudes, in part because it is much easier to figure out which collections of them are consistent or inconsistent. Of course, there is already a study of the consistent sets of logical platitudes for truth—it is the study of axiomatic theories of truth. Hence, we are taking axiomatic theories of truth to be our guide in the study of the alethic properties. The alethic properties will end up being those properties that satisfy the maximal consistent subsets of the logical platitudes for truth. I invite you to think about it this way: although we will find no property even remotely close to what we think of as the property of being true, we will find lots of properties that obey certain subsets of the logical platitudes for truth.<sup>19</sup>

## 4 The Friedman–Sheard Criteria

So far, our Canberra plan for truth is humming along nicely. We have collected a ton of logical platitudes for truth, and we are ready to discover which property uniquely satisfies them. So far we have no official reason to think there isn't a unique property of being true, although I have mentioned that a swarm of paradoxes is soon going to obliterate our plan.

Enter the liar paradox. From (T-In) (i.e., if  $p$ , then  $\langle p \rangle$  is true) and (T-Out) (i.e., if  $\langle p \rangle$  is true, then  $p$ )—one can derive that a sentence like

(1) (1) is not true

is both true and not true. Often the paradox is formulated in terms of falsity, but this is not necessary, and since falsity is often defined in terms of truth, it is a needless

<sup>18</sup> Axiomatic theories of truth constitute one type of formal theory that is formulated using the resources of proof theory; see Halbach (2011) for a survey. There is another tradition that utilizes the tools from model theory instead. These are often called *semantic theories of truth*. See Field (2008) for an overview of this tradition. Roughly, an axiomatic theory of truth specifies a set of formulas involving a truth predicate—the axioms—and rules for deriving new formulas from existing ones; the derived formulas are the theorems. By contrast, a semantic theory of truth specifies a set of true sentences of the object language in question. This set might be thought of as the extension of 'true' or as the set of sentences defining 'true'. There has been a considerable amount of work on relations between axiomatic theories of truth and semantic theories of truth; see Fischer et al (2015). For example, the axiomatic theory, KF, which we consider in Section Seven, is designed to have as theorems all the formulas that are valid in the Strong Kleene version of Kripke's semantic theory. Instead of letting the axiomatic theories be our guide to the alethic properties, we could focus on the semantic theories of truth, but this sort of project would need some other methodology and some other motivation. Our focus on axiomatic theories of truth stems from thinking about logical platitudes for truth within a Canberra plan methodology, and the need to consider consistent subsets thereof.

<sup>19</sup> I am unaware of any theorist who focuses exclusively on the logical platitudes for truth, but there are plenty of discussions of technical principles that might serve as norms for a theory of truth. See Sheard (1994, 2002), Leitgeb (2007), Horsten (2011), and Halbach and Horsten (2015).

distraction.<sup>20</sup> Moreover, the argument relies on the claim that, because sentence (1) is identical to the sentence ‘(1) is not true’, these expressions are intersubstitutable in the truth predicate; this principle is labeled (Sub) in the argument.

- |                                    |                                      |
|------------------------------------|--------------------------------------|
| 1. (1) is true                     | [assumption for reductio]            |
| 2. ‘(1) is not true’ is true       | [(Sub) from 1]                       |
| 3. (1) is not true                 | [(T-Out) from 2]                     |
| 4. (1) is true and (1) is not true | [conjunction introduction from 1, 3] |
| 5. (1) is not true                 | [reductio from 1–4]                  |
| 6. ‘(1) is not true’ is true       | [(T-In) from 5]                      |
| 7. (1) is true                     | [(Sub) from 6]                       |
| 8. (1) is true and (1) is not true | [conjunction introduction from 5, 7] |

There are other ways to formalize the informal liar reasoning that occurs in natural language, but this one is fine for our purposes.<sup>21</sup> However one formalizes the liar reasoning, the fact that one can derive a contradiction in this way from (T-In) and (T-Out) shows that they cannot be satisfied together in a setting that allows classical logic (or any of its familiar alternatives—e.g., intuitionistic logic and the logic of relevant implication).

Why accept the classical inferences used in this derivation of the paradox? We are investigating which alethic properties exist. Because it is commonplace in discussions of *properties* to permit the intuitive ways of reasoning that are counted as valid by classical logic and any of its familiar alternatives, we also take for granted that these ways of reasoning are valid. Surely there is a place in the study of alethic properties for investigations into which properties might exist if one severely restricts these ways of reasoning (as in any of the non-classical approaches to the liar and other paradoxes of truth), but that won’t be our strategy here.<sup>22</sup> Hence, every alethic property will satisfy at most one of (T-In) and (T-Out). This result is essentially Tarski’s theorem on the indefinability of truth transferred into the study of alethic properties.<sup>23</sup>

There are other well-known paradoxes associated with truth, like the Curry paradox and the Yablo paradox. The Curry pertains to a conditional like:

- (2) If (2) is true, then  $0=1$ .

From (T-In) and (T-Out) one can derive that  $0=1$  using (2). By varying the example, one can derive anything at all.

<sup>20</sup> Two common definitions are:

⟨p⟩ is false iff ⟨p⟩ is not true, and  
 ⟨p⟩ is false iff ⟨¬p⟩ is true.

<sup>21</sup> See Scharp (2013) for discussion.

<sup>22</sup> See Field (2004) and Field et al (2017) for examples of non-classical theories of properties.

<sup>23</sup> Tarski (1933); see Field (2008, ch. 1) for discussion.

The Yablo paradox is based on an infinite sequence of sentences each saying of all the ones beyond it that they are not true. Again, from (T-In) and (T-Out), one can derive that they are true and not true.<sup>24</sup> Because the Curry paradox and Yablo paradox involve exactly the same logical platitudes as the liar paradox—namely (T-In) and (T-Out)—these paradoxes won't concern us any more in our investigation. There is no accepted way of individuating paradoxes. However, for the purposes of identifying the alethic properties, we care only about distinguishing paradoxes of truth that involve distinct sets of logical platitudes beyond (T-In) and (T-Out).

It is easy to see that the combination of (T-Intro) and (T-Elim) is inconsistent as well, since they can be used in place of (T-In) and (T-Out), respectively, in the liar reasoning. However, one cannot derive a contradiction from (T-Enter) and (T-Exit) since they cannot be deployed in hypothetical reasoning. In the proof given above, steps 1–4 occur as part of hypothetical reasoning, so we cannot use (T-Exit) in place of (T-Out) in step 3. Hence, some alethic properties do satisfy these two platitudes—we will have to keep them in mind.

Still, this leaves us without much to go on. What we need is some guide to the consistent and inconsistent sets among the logical platitudes for truth. Luckily, Harvey Friedman and Michael Sheard undertook just such an investigation.<sup>25</sup>

They focused on twelve logical platitudes for truth. We call them the *Friedman–Sheard Criteria*:

- (T-In)  $p \rightarrow T\langle p \rangle$
- (T-Out)  $T\langle p \rangle \rightarrow p$
- (T-Enter)  $\vdash p \rightarrow \vdash T\langle p \rangle$
- (T-Exit)  $\vdash T\langle p \rangle \rightarrow \vdash p$
- ( $\neg$ T-Enter)  $\vdash \neg p \rightarrow \vdash T\langle \neg p \rangle$
- ( $\neg$ T-Exit)  $\vdash T\langle \neg p \rangle \rightarrow \vdash \neg p$
- (T-Rep)  $T\langle p \rangle \rightarrow T\langle T\langle p \rangle \rangle$
- (T-Del)  $T\langle T\langle p \rangle \rangle \rightarrow T\langle p \rangle$
- (T-Comp)  $T\langle p \rangle \vee T\langle \neg p \rangle$  [logically equivalent to ( $\neg$ -Imb):  $\neg T\langle p \rangle \rightarrow T\langle \neg p \rangle$ ]
- (T-Cons)  $\neg(T\langle p \rangle \wedge T\langle \neg p \rangle)$  [logically equivalent to ( $\neg$ -Exc):  $T\langle \neg p \rangle \rightarrow \neg T\langle p \rangle$ ]
- (U-Imb)  $(\forall x)T\langle \phi(x) \rangle \rightarrow T\langle (\forall x)\phi(x) \rangle$
- (E-Exc)  $T\langle (\exists x)\phi(x) \rangle \rightarrow (\exists x)T\langle \phi(x) \rangle$

Friedman and Sheard worked meticulously to find every inconsistent subset of these principles, and they gave consistency proofs for each consistent subset.

We can see some of their results in Fig. 1. The Friedman–Sheard criteria are listed on the left and the right sides of the figure, and every subset of platitudes connected by dotted lines is inconsistent. In each case there is a paradox, like the liar paradox, but whose reasoning involves those particular principles. One can see seventeen

<sup>24</sup> There are a number of subtleties in the formulation Yablo's paradox pertaining to omega-inconsistency and inconsistency that do not concern us. See Ketland (2005) and Cook (2014).

<sup>25</sup> Friedman and Sheard (1987). See also Friedman and Sheard (1988), Leigh and Rathjen (2010), and Leigh (2015).

paradoxes depicted as dotted lines (if three or more platitudes are involved, then the dotted lines are connected by a small dot). Each dot and each line with no dots connecting platitudes is a distinct paradox. There are two entailments listed (there are others as well). Because it entails (T-Cons), ( $\neg$ T-Enter) has three additional paradoxes associated with it that are not listed. And because it entails (T-Comp), (E-Exc) has an additional seven paradoxes associated with it that aren't listed. So the two solid lines represent ten paradoxes between them. A grand total of *twenty-seven* (!) distinct paradoxes are lurking in these twelve logical platitudes for truth.

We have to be careful because Friedman and Sheard make several background assumptions in their reasoning:

- (MP)  $(T\langle p \rightarrow q \rangle \wedge T\langle p \rangle) \rightarrow T\langle q \rangle$  [equivalent to ( $\rightarrow$ -Exc):  $T\langle p \rightarrow q \rangle \rightarrow (T\langle p \rangle \rightarrow T\langle q \rangle)$ ]  
 (Taut)  $T\langle p \rangle$  if  $\langle p \rangle$  is a tautology  
 (PRE)  $T\langle p \rangle$  if  $\langle p \rangle$  is an axiom of PRE.

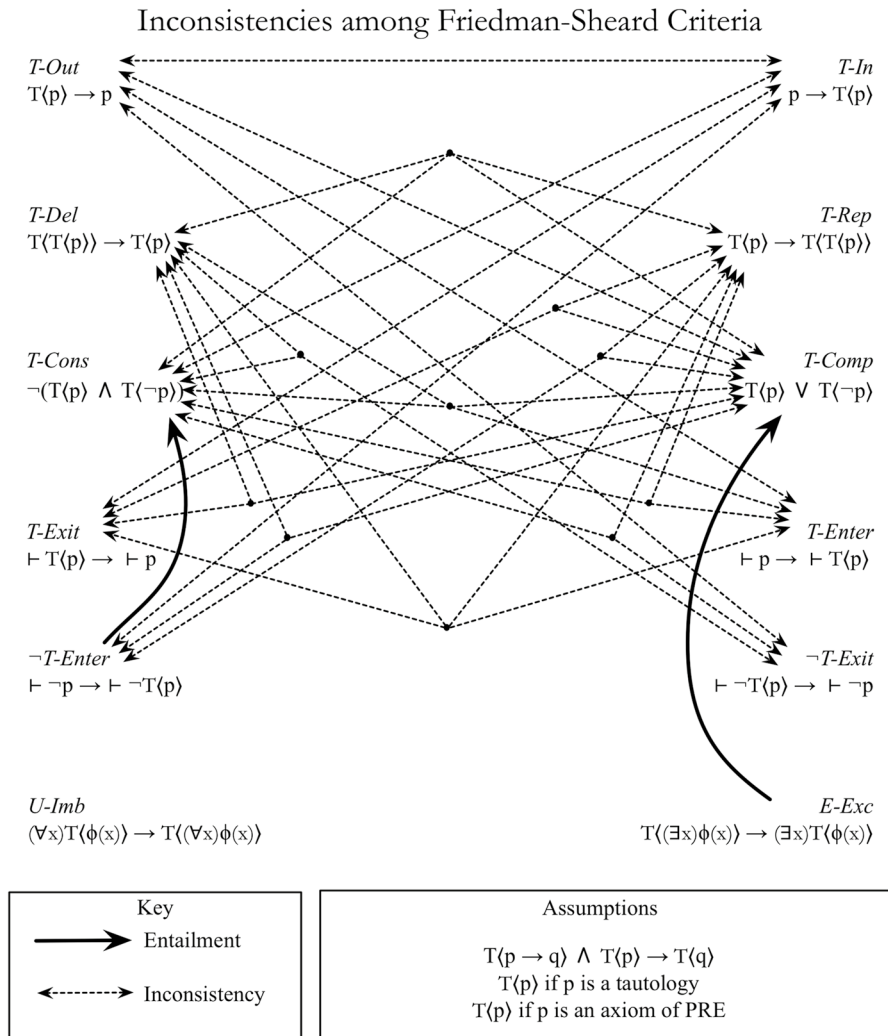
PRE is a mathematical theory that Friedman and Sheard consider as part of their background assumptions. It contains certain mathematical equations. Friedman and Sheard also assume Peano arithmetic, which is a popular axiomatic theory of arithmetic (e.g., a theory of the natural numbers, addition, multiplication, etc.).

In assuming (PRE) and Peano Arithmetic, Friedman and Sheard are following standard protocol in the literature on axiomatic theories of truth. They use numerals in their object language to refer to symbols, expressions and sentences of the object language via a famous method called Gödel numbering. The truth predicate of their object language has only numbers in its extension, and these are intended to be interpreted as the Gödel numbers of the true sentences of the object language. Because Friedman and Sheard use Gödel numbers and arithmetic to prove various things about the sentences of their object language, they need some kind of mathematical theory for these proofs; Peano Arithmetic and PRE are the mathematical theories they use.<sup>26</sup>

Because we care about natural language, we are not concerned with any of these details beyond noting that we achieve the same sorts of aims by assuming various things about sentences of English (e.g., that liar sentences exist and that the liar is identical to the sentence 'the liar is not true'). Given that (MP), (TAUT), and (PRE) are assumed throughout their discussion, it makes sense for us to add them to the list of Friedman Sheard criteria. There is no reason to think that they are somehow more sacrosanct than any of the other platitudes. However, giving up (PRE) would not help much with the aletheic paradoxes in natural language because reliance on (PRE) is largely an artifact of the object language in Friedman and Sheard's treatment. When reasoning about natural language paradoxes we use various assumptions about which words refer to which sentences instead. One of these was listed as (Sub) in the liar reasoning at the beginning of this section.

<sup>26</sup> Friedman and Sheard also use (PRE) in results about the proof theoretic strength of various theories, but this topic is independent of ours.





**Fig. 1** Inconsistencies among the Friedman–Sheard criteria

Although it might be hard to discern from Fig. 1, there are exactly nine maximal consistent subsets of the Friedman–Sheard criteria. They are:

- A.  $T\text{-In}$ ,  $T\text{-Enter}$ ,  $\neg T\text{-Exit}$ ,  $T\text{-Del}$ ,  $T\text{-Rep}$ ,  $T\text{-Comp}$ ,  $U\text{-Imb}$ ,  $E\text{-Exc}$ .
- B.  $T\text{-Rep}$ ,  $T\text{-Cons}$ ,  $T\text{-Comp}$ ,  $U\text{-Imb}$ ,  $E\text{-Exc}$ .
- C.  $T\text{-Del}$ ,  $T\text{-Cons}$ ,  $T\text{-Comp}$ ,  $U\text{-Imb}$ ,  $E\text{-Exc}$ .
- D.  $T\text{-Enter}$ ,  $T\text{-Exit}$ ,  $\neg T\text{-Enter}$ ,  $\neg T\text{-Exit}$ ,  $T\text{-Cons}$ ,  $T\text{-Comp}$ ,  $U\text{-Imb}$ ,  $E\text{-Exc}$ .
- E.  $T\text{-Enter}$ ,  $T\text{-Exit}$ ,  $\neg T\text{-Enter}$ ,  $T\text{-Del}$ ,  $T\text{-Cons}$ ,  $U\text{-Imb}$ .
- F.  $T\text{-Enter}$ ,  $T\text{-Exit}$ ,  $\neg T\text{-Enter}$ ,  $T\text{-Del}$ ,  $U\text{-Imb}$ .
- G.  $T\text{-Enter}$ ,  $T\text{-Exit}$ ,  $\neg T\text{-Exit}$ ,  $T\text{-Rep}$ ,  $U\text{-Imb}$ .

- H. T-Out, T-Exit,  $\neg$ T-Enter, T-Rep, T-Del, T-Cons, U-Imb.  
 I. T-Exit,  $\neg$ T-Exit, T-Rep, T-Del, U-Imb.

In Fig. 2, there is a table of the Friedman–Sheard criteria on the left and the consistent subsets on the top, with ‘✓’ indicating inclusion of the platitude on the left of that row in the subset listed at the top of that column.

Remember that the three additional platitudes that Friedman and Sheard used as background reasoning should be appended to each of the nine consistent subsets of Friedman–Sheard criteria.

Some observations about the table deserve comment.

First, Friedman and Sheard label the maximally consistent subsets in the order in which they prove them consistent, and they begin with the easiest cases first. Other than that there is nothing significant about the order.

Next, Universal Imbibe (U-Imb) is a member of every single consistent subset, and so it is not involved in any of the paradoxes Friedman and Sheard find; none of the other Friedman–Sheard criteria is like this. I shall use the term ‘innocuous’ for this feature, and ‘destructive’ for its converse. With this terminology, we can say that (U-Imb) is *the most innocuous* of the Friedman–Sheard criteria, given their background assumptions.

Note how destructive (T-In) and (T-Out) are individually; that is, each one is inconsistent with lots of other subsets of the Friedman–Sheard criteria. Each one shows up only one time among the consistent subsets of Friedman–Sheard Criteria; (T-In) is in subset A and (T-Out) is in subset H. None of the other criteria are even close to that destructive. Remember, that (T-In) and (T-Out) are the two conditionals conjoined in the biconditional (Schema T). So *each direction* of (Schema T), *all by itself*, is ridiculously destructive. No wonder they cause so much trouble together!

Only one subset contains (T-Enter), (T-Exit), ( $\neg$ T-Enter), and ( $\neg$ T-Exit) together. If one is looking for a more innocuous version of (Schema T) that is consistent, then this combination is one option.<sup>27</sup> Together these four rules say something like: a theory proves a sentence if and only if the theory proves the truth attribution to that sentence, and a theory proves a negated sentence if and only if the theory proves the negation of the truth attribution to that sentence. Again, these rules cannot be used in hypothetical reasoning, as in when one argues using a reductio or conditional proof. Even though this combination of the four derivation rules is considerably more innocuous than (Schema T), they are still *very* destructive together—they occur together only in subset D.

The most important thing to note is how *empty* the table is. Only 54 of the 108 squares are occupied. *Half!* Even if the table were only 75% occupied, that would be really bad—there would be lots of contradictions within those twelve Friedman–Sheard criteria.

The Friedman–Sheard criteria seem to give us *nine* alethic properties—one property for each consistent subset of the Friedman–Sheard criteria. However, it would

<sup>27</sup> See Greenough (2001) for a philosophical motivation for this combination.

	A	B	C	D	E	F	G	H	I
$\begin{array}{c} \text{(T-In)} \\ p \rightarrow T\langle p \rangle \end{array}$	✓								
$\begin{array}{c} \text{(T-Out)} \\ T\langle p \rangle \rightarrow p \end{array}$								✓	
$\begin{array}{c} \text{(T-Enter)} \\ \vdash p \rightarrow \vdash T\langle p \rangle \end{array}$	✓			✓	✓	✓	✓		
$\begin{array}{c} \text{(T-Exit)} \\ \vdash T\langle p \rangle \rightarrow \vdash p \end{array}$				✓	✓	✓	✓	✓	✓
$\begin{array}{c} \text{(\neg T-Enter)} \\ \vdash \neg p \rightarrow \vdash \neg T\langle p \rangle \end{array}$				✓	✓			✓	
$\begin{array}{c} \text{(\neg T-Exit)} \\ \vdash \neg T\langle p \rangle \rightarrow \vdash \neg p \end{array}$	✓			✓		✓	✓		✓
$\begin{array}{c} \text{(T-Comp)} \\ T\langle p \rangle \vee T\langle \neg p \rangle \end{array}$	✓	✓	✓	✓					
$\begin{array}{c} \text{(T-Cons)} \\ \neg T\langle p \rangle \wedge T\langle \neg p \rangle \end{array}$		✓	✓	✓	✓			✓	
$\begin{array}{c} \text{(T-Del)} \\ T\langle T\langle p \rangle \rangle \rightarrow T\langle p \rangle \end{array}$	✓		✓		✓	✓		✓	✓
$\begin{array}{c} \text{(T-Rep)} \\ T\langle p \rangle \rightarrow T\langle T\langle p \rangle \rangle \end{array}$	✓	✓					✓	✓	✓
$\begin{array}{c} \text{(U-Imb)} \\ \langle \forall x \rangle T\langle \phi(x) \rangle \rightarrow T\langle \langle \forall x \rangle \phi(x) \rangle \end{array}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\begin{array}{c} \text{(E-Exc)} \\ T\langle \langle \exists x \rangle \phi(x) \rangle \rightarrow \langle \exists x \rangle T\langle \phi(x) \rangle \end{array}$	✓	✓	✓	✓					

**Fig. 2** Maximal consistent subsets of Friedman–Sheard criteria

make more sense to think of these as *families* of alethic properties because we can imagine adding other logical platitudes to any of these subsets (each subset is maximal among the Friedman–Sheard criteria—in the sense that one cannot consistently add any more *Friedman–Sheard platitudes* to any of these nine subsets—but they

are certainly not maximal among the logical platitudes in general). Moreover, we will see in Section Six that some of these families of properties do not exist because they place unacceptable conditions on how to interpret our arithmetic vocabulary. In the meantime, we can call any property that satisfies the platitudes in subset A, the property of being FSAtrue. We can call any property that satisfies the platitudes in subset B, the property of being FSBtrue. And so on.

At last we have something to work with when thinking about which property is the property of being true.<sup>28</sup> And the results are grim. We are assuming for now that there is a property of being FSAtrue and a property of being FSBtrue and so on for each of our nine families. But which one of these is the property of being true? Of course, none of them satisfy all the Friedman–Sheard criteria. But more damning is that none of them *even come close* to satisfying all the Friedman–Sheard criteria. The best meet two-thirds of the criteria (subset A and subset D). Maybe you think that is close enough to count as the property of being true, but that view runs into another problem—there are two properties that satisfy eight out of the twelve criteria. So even if the Friedman–Sheard criteria were the only truth platitudes, no property comes close to satisfying all of them, and multiple distinct properties seem to do equally good jobs of satisfying some of them.

And we find more paradoxes when we include more logical platitudes.

## 5 More paradoxes

Friedman and Sheard look at twelve logical platitudes for truth but we saw over forty in Section Three. What one would want at this point in our investigation is a catalog like Friedman and Sheard’s project but for a ton more logical platitudes. Unfortunately, no such thing exists. As a prolegomena, we could consider several glimpses of the world beyond the Friedman–Sheard criteria in the form of paradoxes affecting this broader class of logical platitudes for truth. Several stand out.

Richard Montague in 1963 proved a theorem whose intended interpretation was about necessity.<sup>29</sup> At the time, many philosophers thought that necessity might be best understood as a predicate, ‘x is necessary’, where ‘x’ gets filled in by a name or description, rather than as an operator, ‘it is necessary that p’, where ‘p’ gets filled in by a sentence. However, this result of Montague’s showed that if necessity is a predicate, then it cannot be anything like what we take necessity to be. Since then, the operator treatment of necessity has become standard and the predicate treatment almost non-existent. But we have a different application in mind—truth. Interpreted as a limiting theorem for truth, just as Tarski’s indefinability theorem, Montague’s result says that the following logical platitudes for truth are inconsistent:

<sup>28</sup> An alternative approach would investigate properties that satisfy *many but not all* instances of certain truth platitudes rather than properties that satisfy *all* the truth platitudes in some maximal consistent set of platitudes. For example, Horwich (1998) makes a suggestion like this for just (Schema T). Evaluating this alternative is beyond the scope of this paper. See Horwich (1998), McGee (1992), and Schindler (2015) for discussion.

<sup>29</sup> Montague (1963).

(T-Out)	$T\langle p \rangle \rightarrow p$
(T-TOut)	$T\langle T\langle p \rangle \rightarrow p \rangle$
(MP)	$(T\langle p \rightarrow q \rangle \wedge T\langle p \rangle) \rightarrow T\langle q \rangle$ [which is logically equivalent to $((\rightarrow -\text{Exc})]$
(Taut)	$T\langle p \rangle$ if $\langle p \rangle$ is a tautology
(Robinson)	$T\langle p \rangle$ if $\langle p \rangle$ is a truth of Robinson arithmetic

(T-Out) is one of the Friedman–Sheard criteria, while (MP) and (Taut) are familiar from the Friedman–Sheard background assumptions as well. Note that Montague’s theorem assumes the truth of the theorems of Robinson arithmetic instead of PRE, like in Friedman and Sheard. Robinson Arithmetic is a subtheory of Peano Arithmetic, which is also assumed as part of the background by Friedman and Sheard, and it is present in Montague’s proof for the same reason (to prove various things about the sentences of the object language).

The new kid is (T-TOut), which says that (T-Out) is true. Note that in the case of necessity, Montague’s theorem is less threatening because (T-TOut) would be the claim that ‘if  $p$  is necessary then  $p$ ’ is necessary. Denying that an axiom of the proper theory of necessity is necessary is somewhat awkward, but not nearly as bad as a self-refutation, which occurs in the case of truth. That is, to deny (T-TOut) but accept (T-Out) for truth is to take one’s own theory of truth to be not true. Therefore, Montague’s theorem is even more troublesome in the realm of truth than it was in the realm of necessity, where it laid waste to the assumptions underlying generations of research on modality.

When we translate Montague’s theorem into talk of alethic properties, we get something like: for any alethic property, either it isn’t preserved by central inference rules like modus ponens, it isn’t had by all logical tautologies and mathematical theorems, it isn’t a guide to the world (e.g., a sentence  $p$  can have the alethic property even though not- $p$ ), or it isn’t had by the claim that it is a guide to the world. *Every alethic property has one of these highly undesirable features.*

Montague’s stunning result can be brought into fruitful interaction with the Friedman–Sheard criteria: the three background conditions for the Friedman–Sheard criteria together entail two of the four interesting logical platitudes shown to be inconsistent by Montague’s theorem. And (T-Out), one of the assumptions of Montague’s theorem, is another of the Friedman–Sheard criteria. (T-TOut) is the only assumption of Montague’s theorem that is absent from the framework constructed by Friedman and Sheard (other than (Robinson)).

Montague’s theorem shows that one cannot consistently add (T-Out) and (T-TOut) to any of the consistent subsets of Friedman–Sheard criteria. Recall that (T-Out) is a member of only one of the Friedman–Sheard families—subset H. So we know that we cannot add (T-TOut) to subset H. Any theory like H that includes (T-Out) is going to prove a liar sentence and prove that the liar sentence is not true. So any theory that includes (T-Out) proves that some of its own theorems are not true (more on this issue below in section nine). However, when we factor in Montague’s theorem, we see that theories like H cannot even be said to have all true *axioms*, much less all true theorems.

The next result is not prominent in the literature on truth, but it is based on Gödel's second incompleteness theorem. Roughly, Gödel's second says that certain formal systems cannot prove their own consistency if they are consistent. Or, formal systems in a certain class can prove their own consistency only if they are inconsistent. The theorem includes several assumptions about proof, and it defines consistency in terms of proof as well. However, if one reinterprets these provability criteria and the definition of consistency as logical platitudes about *truth*, then one proves that the following platitudes are inconsistent:

- (T-Enter)  $\vdash p \rightarrow \vdash T\langle p \rangle$
- (T-Rep)  $T\langle p \rangle \rightarrow T\langle T\langle p \rangle \rangle$
- (MP)  $(T\langle p \rightarrow q \rangle \wedge T\langle p \rangle) \rightarrow T\langle q \rangle$
- (Contra)  $\neg T\langle p \rangle$  for  $p$  a contradiction
- (Robinson)  $T\langle p \rangle$  for  $p$  a theorem of Robinson Arithmetic

Two of the platitudes are among the Friedman–Sheard criteria—(T-Enter) and (T-Rep)—and one is a background assumption for the Friedman–Sheard framework: (MP). (Contra) is related to (Taut), which is among the Friedman–Sheard background assumptions, but (Contra) does not occur at all in the Friedman–Sheard framework. Only one of their subsets has (T-Enter), (T-Rep), and (MP)—subset G. Thus, by Gödel's second incompleteness theorem, subset G defines an alethic property that holds of contradictions. It is a *dialetheic* property. If the alethic property that is satisfied by subset G of Friedman–Sheard criteria were somehow picked to be the best candidate for the property of being true, then this result would vindicate dialetheism. Moreover, unless everything is true, the internal logic of a dialetheic truth predicate must be very weak—weak enough so that it is impossible to reason from the truth of the contradiction to the truth of an arbitrary sentence via *ex falso*.

It might seem odd that there could be a dialetheic property of being true in the classical framework we are using as a background. Typically, dialetheism is paired with a paraconsistent logic, in which one cannot derive an arbitrary sentence from a contradiction, but there is a version of dialetheism that is consistent with classical logic—it entails that some contradictions are true, but it does not entail  $p$  and not  $p$  for some  $p$ . The reason is that this classical version of dialetheism does not have any way to “excrete” the contradiction from inside the truth predicate; it would need something like (T-Out) or (T-Exit), neither of which are in subset G. As far as I know, no one endorses this classical version of dialetheism—something like subset G; the main reason is that one of the main benefits of dialetheism is thought to be that it allows one to keep (Schema T) and so all the disquotational platitudes, but this benefit comes at the cost of giving up classical logic for paraconsistent logic.<sup>30</sup>

Vann McGee discovered another paradox worth mentioning here.<sup>31</sup> McGee's result says that any theory containing the following platitudes is omega-inconsistent:

<sup>30</sup> See Field (2008: ch. 8) for discussion.

<sup>31</sup> McGee (1991).

(T-Enter)	$\vdash p \rightarrow \vdash T\langle p \rangle$	
(MP)	$(T\langle p \rightarrow q \rangle \wedge T\langle p \rangle) \rightarrow T\langle p \rangle$	[which is logically equivalent to $((\rightarrow -\text{Exc})]$
$(\neg\text{-Exc})$	$T\langle \neg p \rangle \rightarrow \neg T\langle p \rangle$	[which is logically equivalent to $(\text{Consist})]$
(U-Imb)	$(\forall x)T\langle \phi(x) \rangle \rightarrow T\langle (\forall x)\phi(x) \rangle$	
(Robinson)	$T\langle p \rangle$ if $\langle p \rangle$ is a theorem of Robinson Arithmetic	

(MP) is an old friend (enemy?) by now, having shown up in the background assumptions to the Friedman–Sheard results, Montague’s theorem, Gödel’s second theorem, and now McGee’s theorem. (U-Imb), (T-Enter), and  $(\neg\text{-Exc})$  are among the Friedman–Sheard criteria.

Notice, McGee didn’t prove these to be inconsistent, but instead showed that they are *omega*-inconsistent. An omega-inconsistent set of sentences proves some existential generalization (e.g., for some  $x$ ,  $x$  is  $F$ ) and it also proves the negation of each instance (e.g.,  $a$  is not  $F$ ,  $b$  is not  $F$ ,  $c$  is not  $F$ , ...). Omega-inconsistent sets might be consistent, so McGee’s theorem has a weaker conclusion than Montague’s theorem or Gödel’s second theorem (in fact, McGee proves that it cannot be strengthened to an inconsistency result).

There is an important technical result about omega-inconsistency: omega inconsistent theories are inconsistent with the standard interpretation of arithmetical vocabulary. That is, if an omega-inconsistent theory formulated in some language that contains the usual arithmetic vocabulary (e.g., numerals, addition sign, ...), then that arithmetic vocabulary cannot have its intended interpretation. In particular, it turns out that in omega-inconsistent theories formulated in languages with arithmetic vocabulary, the quantifiers do not range over only natural numbers.

An assumption of the discussion so far is that reality is consistent; hence, inconsistent theories do not describe any real properties. Omega-inconsistent theories are consistent, so one might think that they describe real properties, but in fact the opposite conclusion is in order. If reality is consistent and our arithmetic vocabulary really has the obvious interpretation, then omega-inconsistent theories don’t describe real properties at all. Instead, the right view is that real properties are described by theories that do not require wild and radically counterintuitive interpretations of basic arithmetic. Therefore, in the study of alethic properties, we should take McGee’s theorem to be a limiting theorem just like Montague’s theorem and Tarski’s theorem. That is, we should think of McGee’s paradox as on par with the rest.

In their recent evaluation of axiomatic theories of truth, Leon Horsten and Volker Halbach respond to the objection from omega-inconsistency. They write: “Certainly  $\omega$ -consistency is an important requirement, but we do not see it as a very fundamental requirement: hardly anyone would have thought of imposing  $\omega$ -consistency as a requirement before McGee (1985) proved the  $\omega$ -inconsistency of a vast class of truth theories that looked otherwise very attractive.”<sup>32</sup> This is not a good reason for thinking that omega consistency is not a “very fundamental requirement”. The fact

<sup>32</sup> Horsten and Halbach (2015: p. 268) and they are responding to Leitgeb (2007).

that omega-inconsistent theories are not compatible with the standard interpretation of arithmetic should be decisive: there are no properties that satisfy these theories of truth; otherwise, we are seriously misinformed about meanings of the arithmetic bits of our own language. That is not a price anyone should be willing to pay for choosing one relatively arbitrary combination of truth platitudes over one of the other relatively arbitrary combinations. In general, one should try to keep the costs of one's philosophical views about X within the same realm so that collateral damage affecting independent topics is to be avoided. And that is exactly what an omega-inconsistent theory does. The fact that very few logicians happened to realize that some intuitive theories of truth are omega-inconsistent before McGee's result is irrelevant.

McGee's theorem has serious consequences for the consistent subsets of the Friedman–Sheard criteria: any consistent subset of Friedman–Sheard criteria that includes (U-Imb), ( $\neg$ -Exc) and (T-Enter) is omega-inconsistent. Therefore, McGee's theorem shows that subset D and subset E are omega-inconsistent. Given our conclusion that omega-inconsistent sets of truth platitudes do not describe alethic properties, we have a significant negative result—there is no property of being FSDtrue and there is no property of being FSEtrue.

There are so many more limiting theorems based on paradoxes among the logical platitudes for truth, but the case is pretty conclusive at this point.

## 6 The nightmare

There are twenty-seven distinct paradoxes among the twelve Friedman–Sheard criteria, and there are only nine maximal consistent subsets of these criteria. The largest two maximal subsets have eight members, but the average size of the maximal subsets is only six. There are 495 distinct subsets of the Friedman–Sheard criteria with eight members, but only *two* of these subsets are consistent! Let that sink in for a moment.

Moreover, if we were to weight the Friedman–Sheard criteria, (T-In) and (T-Out) would easily get the highest weights because of how central they are to the functioning of 'true'. Recall that these are each direction of the celebrated (Schema T). But each of these central criteria show up in only one of the maximal subsets.

However, the situation is actually much worse. Friedman and Sheard write, "The consistency of [(T-In) and (T-Del)] is easily verified by considering the model in which [every sentence of the object language is true]. While this model may seem distant from our original motivation for T, it certainly works, and moreover we will shortly see that it is the only model at all in which T-In holds."<sup>33</sup> Subset A has only one model (given their assumptions), and in that model, the truth predicate's extension contains every sentence. Hence, in that model, everything is true. That is, all the sentences in the language are in the extension of the truth predicate of the language. (Note that this result is different from having every sentence of the language

<sup>33</sup> Friedman and Sheard (1987: pp. 7–8); note that they are assuming that all models have the natural numbers as their domain and that the arithmetic vocabulary has the standard interpretation.



be true in that model. What the model calls true and what the model calls ‘in the extension of the truth predicate’ need not be the same.) Having assumptions about truth and arithmetic that force all sentences of a language to be in the truth predicate’s extension is so implausible that we can ignore subset A in what follows.

Recall that subset A was the only one that included (T-In). And, the subset of which (T-Out) is a member, subset H, has models that do not interpret the truth predicate of the language as having all the sentences in its extension. In some sense, (T-In) is the more destructive of the two directions, given Friedman and Sheard’s background assumptions.

Moreover, McGee’s theorem proves that subset D and subset E are omega-inconsistent (although they are consistent). We have found that omega-inconsistent sets do not describe any alethic properties.

Therefore, no alethic properties satisfy three—A, D, E—of the nine consistent subsets of Friedman–Sheard criteria. Only Subsets B, C, F, G, H, and I are satisfied by properties at all.

In addition, Subset G is dialethic, by Gödel’s second theorem. That means that some contradictions are in the extension of the truth predicate described by Subset G. Or, in other words, some contradictions are FSGtrue. Note that Subset G commits one to *classical* dialetheism (what Field calls the classical glutty view), which no theorists endorse. It has the worst of both worlds—no (Schema T) and some contradictions are true. Graham Priest, Jc Beall and the other major dialetheists are *paraconsistent* dialetheists.<sup>34</sup> They reject classical logic for paraconsistent logic, which allows them to endorse much stronger logical platitudes for truth than are available for classical logic fans. (All of the Friedman and Sheard discussion assumes classical logic.) Because Subset G has such a violent combination of costs and no benefits we shall not consider it further (Field ignores it too).

Next, Montague’s theorem shows that subset H is self-refuting in the sense that not all members of subset H have the property of being FSHtrue. In particular, T-Out is a member of subset H, but it is not FSHtrue. When we eliminate subsets A, D, E, G, and H, we are left with only *four* subsets of Friedman–Sheard criteria, and each of these subsets contains only *five* members. Five! Out of twelve criteria. And none of these four subsets—B, C, F, and I—contain either of the most important criteria: (T-In) and (T-Out).

The number of ways in which one can derive contradictions from the logical platitudes about truth is simply mind-boggling. This is exactly the nightmare scenario the Canberra plan faithful pray never happens. All the candidate properties are so far from satisfying even minimal subsets of truth platitudes, that there is no decent way to judge which one of the alethic properties is *the* property of being true. They are so far from what we think of when we think of the concept of truth that it does

<sup>34</sup> See Priest (2006) and Beall (2009).

not make sense to call any of them the property of being true. The result of this Canberra plan exercise is that there is no property of being true.<sup>35,36</sup>

So what are the options for our Canberra plan now that it has encountered this kind of catastrophe?

## 7 Options

For anyone tempted to think that the lesson should be pluralism about truth, think again. All the truth properties claimed to exist by the truth pluralist are assumed to satisfy *all* the platitudes regarding the concept of truth.<sup>37</sup> But we already know that no alethic properties satisfy even small numbers of logical platitudes. Pluralism is an option when *too many* properties satisfy all the platitudes.

Moreover, truth pluralists typically take distinct properties of being true to be operative in distinct domains of inquiry (e.g., correspondence in science and coherence in ethics). But there is no reason to think that each of the various alethic properties is specific to some domain of discourse.

One response to a catastrophic Canberra plan failure is an error-theoretic treatment of the topic in question. The heart of any error theory about some topic is the claim that all atomic sentences about that topic are false. For example, moral error theorists take all atomic sentences with occurrences of ‘morally good’ to be false.<sup>38</sup> Likewise, the error theorist about truth would hold that all atomic sentences with occurrences of truth predicates are false.

In the moral case, error theories are radically implausible, but at least they are internally consistent. However, in the case of truth, it does not take much reflection to see that an error theory would be self-refuting. That is, it would be impossible for such a theory to be true since ‘the error theory of truth is true’ is an atomic sentence with an occurrence of ‘true’. Hence, the error theory of truth implies that the error theory of truth is not true. Moreover, all the problems we have seen for satisfying the platitudes for the concept truth carry over to the platitudes for the concept of falsity because of how intimately these two concepts are related. Hence, an error theorist about truth will also be an error theorist about falsity. The result would be that all atomic sentences with occurrences of falsity predicates are false. Among the sentences declared to be false would be, of course, all the pronouncements of the error theory of truth and the error theory of falsity. Therefore, an error theory of

<sup>35</sup> If I *had* to pick, I would follow Jeremy Wyatt who (personal communication) suggested Subset F as the closest to what the property of being true would be like because it alone among the survivors has both directions of some kind of disquotational platitude: (T-Enter) and (T-Exit) – as well as (¬T-Exit).

<sup>36</sup> Note that some deflationists [e.g., expressivists like Schroeder (2010)] might welcome the conclusion that there is no property of being true. Still, disquotationalists and minimalists, which make up the overwhelming majority of deflationists, do accept that there is a property of being true, just not a substantive property. More on this in Section Ten.

<sup>37</sup> See Lynch (2009).

<sup>38</sup> See Mackie (1977) and Olson (2014) for examples.

truth is a non-starter.<sup>39</sup> Such is the problem with using the defective concept of truth to try to characterize what is wrong with the defective concept of truth.

## 8 Conceptual engineering

We need a better methodology—the Canberra plan is not helpful in situations like this one, and it obscures many of our options because it is fundamentally a *static* methodology, it is nothing to do with change or improvement. Indeed, David Lewis, who has done more to promote the Canberra plan than anyone else, writes: “One comes to philosophy already endowed with a stock of opinions. It is not the business of philosophy either to undermine or to justify these preexisting opinions, to any great extent, but only to try to discover ways of expanding them into an orderly system,” (Lewis 1973: p. 88).

Instead, we need a methodology that encourages us to evaluate our concepts and, when they are found to be problematic, to improve our conceptual scheme. This is exactly the position advocated by those in the new philosophical movement called *conceptual engineering*.<sup>40</sup> Conceptual engineering projects are remarkably diverse, and examples include Sally Haslanger’s ameliorative project, in which she defends replacing our everyday concept of *woman* with the concept of a *person oppressed on the basis of stereotypical female characteristics*.<sup>41</sup> Theodore Sider’s suggestion that we replace our existential quantifiers and other elements of natural language with surrogates that are stipulated to be metaphysically fundamental is another kind of conceptual engineering project.<sup>42</sup> Given the crucial roles that truth plays in our lives, and the fact that our Canberra plan analysis of truth reached the conclusion that there simply is no property of being true, a conceptual engineering project for truth is in order.

We know that none of the alethic properties can do what we expect the property of being true to do because none of them even come close to satisfying all the alethic platitudes. And since none of these properties is designated by the word ‘true’, which is assumed to express the concept of truth, we can start thinking about the families of alethic *concepts*, each of which corresponds to one of the alethic *properties*. That is, we should think of each alethic property as having its own alethic concept, distinct from the concept of truth. We can think of each of these concepts as an addition to our conceptual scheme. In the terminology from here on, they will be known as the *new alethic concepts*. They are related to one another in the same way as the alethic properties described above. To provide a foundation for answering the overarching conceptual engineering question of how we ought to

<sup>39</sup> For a contrary view on error theories of truth, see Devlin (2003).

<sup>40</sup> For overviews see Burgess and Plunkett (2013a, b) Cappelen (2018), and Scharp (2020); see also Novaes (2018) and Novaes and Reck (2017) for a way of understanding Carnapian explication that is similar to conceptual engineering.

<sup>41</sup> See Haslanger (2000)

<sup>42</sup> See Sider (2011).

change our conceptual scheme in light of these findings, I propose that we take the alethic properties as a guide to the potential alethic concepts we might add to our conceptual scheme.

## 9 Some alethic properties

We have no idea how many consistent maximal axiomatic theories there are among, say, the forty or so logical platitudes listed so far. We do have several high-profile axiomatic theories, and we can think of each one as describing an alethic property. Because these axiomatic theories incorporate about as many logical platitudes about truth as we know how to do right now, it is natural to think that each of these axiomatic theories describes a fairly precise alethic property.

Friedman and Sheard advocate a slight modification of Subset D, which has come to be known as Theory FS.<sup>43</sup> FS is the only popular axiomatic theory of truth that incorporates the four derivation rules from the Friedman–Sheard criteria: (T-Enter), (T-Exit), ( $\neg$ T-Enter), and ( $\neg$ T-Exit). That is, FS is the only popular axiomatic theory of truth to incorporate what we know to be a consistent alternative to the (Schema T). Unfortunately, FS is omega-inconsistent. Thus, given the standards for investigating alethic properties that we have followed, it does not describe any alethic property and shouldn't be considered further in the study of alethic properties.

Another axiomatic theory of truth is inspired by Saul Kripke's paper "Outline of a Theory of Truth," which developed a framework for thinking about non-classical theories of truth—theories that are inconsistent in classical logic, but consistent in vastly weaker logics. For example, Kripke showed how to accept Schema T without contradiction in a relatively weak logic called  $K_3$ . Influenced by Kripke's work, Solomon Feferman formulated an axiomatic theory in a classical framework that is intended to model the class of truths Kripke constructed in a non-classical logic. The result is an axiomatic theory called KF:

- |        |  |                             |
|--------|--|-----------------------------|
| (KF1)  | $T\langle p \rangle \leftrightarrow p$   | for p atomic and arithmetic |
| (KF2)  | $T\langle \neg p \rangle \leftrightarrow \neg p$   | for p atomic and arithmetic |
| (KF3)  | $T\langle \neg \neg p \rangle \leftrightarrow T\langle p \rangle$  |                             |
| (KF4)  | $T\langle p \wedge q \rangle \leftrightarrow T\langle p \rangle \wedge T\langle q \rangle$                                     |                             |
| (KF5)  | $T\langle \neg(p \wedge q) \rangle \leftrightarrow T\langle \neg p \rangle \vee T\langle \neg q \rangle$                       |                             |
| (KF6)  | $T\langle p \vee q \rangle \leftrightarrow T\langle p \rangle \vee T\langle q \rangle$   |                             |
| (KF7)  | $T\langle \neg(p \vee q) \rangle \leftrightarrow T\langle \neg p \rangle \wedge T\langle \neg q \rangle$                       |                             |
| (KF8)  | $T\langle p \rightarrow q \rangle \leftrightarrow (T\langle \neg p \rangle \vee T\langle q \rangle)$                           |                             |
| (KF9)  | $T\langle \neg(p \rightarrow q) \rangle \leftrightarrow (T\langle p \rangle \wedge T\langle \neg q \rangle)$                   |                             |
| (KF10) | $T\langle p \leftrightarrow q \rangle \leftrightarrow (T\langle p \rangle \leftrightarrow T\langle q \rangle)$                 |                             |
| (KF11) | $T\langle \neg(p \leftrightarrow q) \rangle \leftrightarrow (T\langle \neg p \rangle \leftrightarrow T\langle \neg q \rangle)$ |                             |

<sup>43</sup> Warning: authors publishing in this area often mean slightly different things when they talk about the Friedman–Sheard theory; and the same goes for all of the other commonly discussed axiomatic theories.

- (KF12)  $T\langle(\forall x)\phi(x)\rangle \leftrightarrow (\forall x)T\langle\phi(x)\rangle$   
 (KF13)  $T\langle\neg(\forall x)\phi(x)\rangle \leftrightarrow (\exists x)T\langle\neg\phi(x)\rangle$   
 (KF14)  $T\langle(\exists x)\phi(x)\rangle \leftrightarrow (\exists x)T\langle\phi(x)\rangle$   
 (KF15)  $T\langle\neg(\exists x)\phi(x)\rangle \leftrightarrow (\forall x)T\langle\neg\phi(x)\rangle$   
 (KF16)  $T\langle p \rangle \leftrightarrow T\langle T\langle p \rangle \rangle$   
 (KF17)  $T\langle\neg p \rangle \leftrightarrow T\langle\neg T\langle p \rangle \rangle$

Whereas FS tries to incorporate a symmetric grouping of surrogates for (T-In) and (T-Out)—in the form of (T-Enter), (T-Exit), ( $\neg$ T-Enter), and ( $\neg$ T-Exit)—KF avoids all of these platitudes and instead focuses on interactions between the truth predicate and logical vocabulary. However, it does not contain either of the negation rules, which is why it has a clause for each kind of connective along with a clause for each kind of connective negated. One can consistently add either of the negation rules, ( $\neg$ -Imb) or ( $\neg$ -Exc) to KF, but not both. Hence, one could add either (T-Cons) or (T-Comp) to KF, but not both.

KF is consistent and omega-consistent, so there is an alethic property, call it being KFtrue, that satisfies all of the axioms of KF. Being KFtrue is still very far from what we think of as the property of being true because being KFtrue fails to satisfy all sorts of logical platitudes for truth, especially the most important ones. All the general platitudes related to the (Schema T) fail—that is (T-In), (T-Out), (T-Enter), (T-Exit), ( $\neg$ T-Enter), and ( $\neg$ T-Exit) all fail in KF. (Only the platitudes specific to truth, (T-Rep) and (T-Del), are axioms of KF.) So, it is consistent with KF that  $p$  but  $\langle p \rangle$  is not true, and that  $\langle p \rangle$  is true but that not  $p$ .<sup>44</sup>

How does KF relate to the Friedman–Sheard criteria? KF has both (T-Rep) and (T-Del). Only three Friedman–Sheard subsets have these two axioms: A, H, and I. However, subset A has (T-In), but KF doesn't, so KF isn't an extension of subset A. H has (T-Out), so KF isn't an extension of subset H. And I has (T-Exit), so KF isn't an extension of subset I. Instead, KF is not an extension of any of the Friedman–Sheard subsets. Indeed, it violates (Taut), which is one of the Friedman–Sheard background assumptions. That is, some instances of logical tautologies are not KFtrue.

Another influential axiomatic theory, first proposed by Andrea Cantini, is also inspired by Kripke's work, but this time by the supervaluational version of Kripke's approach, rather than the strong Kleene version, which is the basis for KF.

- (VF1)  $p \rightarrow T\langle p \rangle$ , for  $p$  atomic and arithmetic, and for  $p$  negated atomic and arithmetic.  
 (VF2)  $T\langle p \rangle \rightarrow p$   
 (VF3)  $T\langle p \rangle \rightarrow T\langle T\langle p \rangle \rangle$   
 (VF4)  $T\langle\neg T\langle p \rangle \rangle \rightarrow T\langle\neg p \rangle$   
 (VF5)  $\neg(T\langle p \rangle \wedge T\langle\neg p \rangle)$   
 (VF6)  $(\forall x)T\langle\phi(x)\rangle \rightarrow T\langle(\forall x)\phi(x)\rangle$

<sup>44</sup> (T-Out), (T-Enter), and (T-Exit) can each be added consistently to KF, so some of the specific examples can be avoided but the main point – that being KFtrue is not the property of being true – stands.

(VF7)  $(T\langle p \rightarrow q \rangle \wedge T\langle p \rangle) \rightarrow T\langle q \rangle$

(VF8)  $T\langle p \rangle$  if  $\langle p \rangle$  is a tautology

This theory is called VF in honor of Bas van Fraassen, who popularized the supervaluation approach just prior to Kripke's work. Notice that the first two axioms of VF are background assumptions for the Friedman–Sheard Criteria. It turns out that VF is closely related to subset H of the Friedman–Sheard criteria. The difference is that H has (T-Del) whereas VF has something like (T-Del) but with a negation, and H has ( $\neg$ T-Enter) whereas VF does not. We can call the property that satisfies VF being VFtrue.

One interesting feature of VF is that it takes (T-Out) as an axiom. A consequence is that it declares some of its own consequences to be not VFtrue. For example, consider something like a liar sentence for VFtruth:

(3)

(3) is not VFtrue.

Using (T-Out), we can derive that (3) is not VFtrue using the first half of the liar reasoning outlined above. Notice, however, that this result just is sentence (3). So VF entails sentence (3) and it entails that sentence (3) is not VFtrue. Thus, although VF is a consistent theory, it is self-refuting in the sense that it entails that some of its own theorems are not VFtrue.<sup>45</sup> So, of the three axiomatic theories so far, FS, KF, and VF, one is omega-inconsistent and one is self-refuting. Not a good start.

Consider two additional alethic properties, being ascending true and being descending true. These properties are denoted by the concepts of ascending truth and descending truth. Elsewhere I advocate using them as a team of replacements for the concept of truth.<sup>46</sup> Here I am only concerned with presenting them alongside the other alethic properties, and presenting the axiomatic theory of ascending truth and descending truth (ADT) along side the other axiomatic theories.

These two replacements are designed to split the two directions of the (Schema T), which we have been calling (T-In) and (T-Out). One concept, ascending truth, is such that if  $p$  then  $\langle p \rangle$  is ascending true. The other, descending truth, is such that if  $\langle p \rangle$  is descending true, then  $p$ . The converse rules fail in general, but are valid for a certain class, which are called the *safe* sentences. That is, if  $\langle p \rangle$  is safe and  $\langle p \rangle$  is ascending true, then  $p$ ; if  $\langle p \rangle$  is safe and  $p$ , then  $\langle p \rangle$  is descending true. There are a host of other principles for each concept as well. In the following,  $A(x)$  (for ' $x$  is ascending true'),  $D(x)$  (for ' $x$  is descending true'), and  $S(x)$  (for ' $x$  is safe') are all monadic, univocal, and invariant.

(D1)  $D\langle p \rangle \rightarrow p$

<sup>45</sup> KF was criticized for not having (T-Out) because it makes being KFtrue dissimilar from what we think the property of being true should be like. But here VF is criticized for having (T-Out) because it makes VF self-refuting. Such is the way with consistent models of inconsistent concepts.

<sup>46</sup> Scharp (2013).

- (D2)  $D(\neg p) \rightarrow \neg D(p)$
- (D3)  $D(p \wedge q) \rightarrow (D(p) \wedge D(q))$
- (D4)  $(D(p) \vee D(q)) \rightarrow D(p \vee q)$
- (D5)  $D(p)$  if  $p$  is a logical tautology
- (D6)  $D(p)$  if  $p$  is a theorem of Peano Arithmetic
- (D7)  $D(p)$  if  $p$  is an axiom of ADT
- (A1)  $p \rightarrow A(p)$
- (A2)  $\neg A(p) \rightarrow A(\neg p)$
- (A3)  $(A(p) \vee A(q)) \rightarrow A(p \vee q)$
- (A4)  $A(p \wedge q) \rightarrow (A(p) \wedge A(q))$
- (A5)  $\neg A(p)$  if  $p$  is a contradiction
- (A6)  $\neg A(p)$  if  $\neg p$  is a theorem of PA
- (M1)  $D(p) \leftrightarrow \neg A(\neg p)$
- (M2)  $S(p) \leftrightarrow (D(p) \vee \neg A(p))$
- (M3)  $(p \wedge S(p)) \rightarrow D(p)$
- (M4)  $(A(p) \wedge S(p)) \rightarrow p$

Some of the axioms are redundant. Further, these axioms are not meant to be *the* theory of ascending truth and descending truth; rather any theory of ascending truth and descending truth should have the preceding as a subtheory.

Consider two examples involving ‘ascending true’ and ‘descending true’ in sentences that are similar to liars and might be thought to make trouble:

- (4) (4) is not ascending true.
- (5) (5) is not descending true.

We can prove that (4) and (5) are each ascending true and not descending true using the same kind of reasoning as in the liar paradox. However, this just shows that they are unsafe—there is no contradiction here.

It might seem that ADT is self-refuting like the theory VF. Properties satisfying self-refuting theories are genuine properties. But they are very far from what we think of as the property of being true. Moreover, the fact that they are self-refuting is a serious cost when it comes to evaluating conceptual engineering projects for truth. However, ADT is not self-refuting because all the axioms of ADT are descending true (and so ascending true).

Still, one might object that some *theorems* of ADT are not descending true, so it is self-refuting after all, just in a weaker sense. Indeed, ADT does imply that some of its own theorems are not descending true. For example ‘(10) is not descending true’ is a theorem of ADT and ADT implies that (10) is not descending true. So ADT implies that some of its theorems are unsafe. (Incidentally, this feature of ADT is why there are redundancies in its formulation. Axiom schema D7 insures that all its axioms are descending true, but one is not thereby guaranteed that all its theorems are descending true. The more axioms one lists, the more are guaranteed to be descending true.)

There are two points to be made in reply. First, if one has only one alethic status to work with, e.g., VFtruth, then the self-refutation considerations given above are decisive. However, if one has two alethic notions to work with, like ascending truth and descending truth, then we can formulate a criterion of adequacy for good (i.e., trustworthy) arguments that the theory, ADT, respects: namely a valid argument will never take one from descending truths to something not ascending true. It might take one from descending truth to unsafety or from unsafety to something not ascending true. For example, the descending liar—(5)—is provable in ADT, the axioms of ADT are descending true, but the descending liar is unsafe (so not descending true). Also, the ascending liar and its negation are unsafe (so ascending true), but their conjunction is a contradiction and so is not ascending true. Second, anyone who accepts that the new alethic statuses are not preserved by valid arguments will be forced to distinguish between theorems of the theory that have top status, those that have middle status(es), and those that have bottom status. If valid arguments never take one from top to bottom, then the obvious condition on an acceptable theory is that all its axioms have top status. We know already that not all its theorems will have top status. So the best we can hope for is top status axioms and no bottom status theorems. That is exactly the case with ADT: all the axioms of the theory are descending true and none of its theorems are not ascending true. So it behaves exactly as one would hope a theory of this type would behave.

A summary is in order. We began with the twelve Friedman Sheard criteria and the 27 inconsistencies among them. Only four of the maximally consistent subsets are even worth considering as being satisfied by anything like what we think the property of being true should be like. Although there is no property of being true, these four are our first good examples of alethic properties. We looked beyond the Friedman–Sheard criteria to other logical platitudes for truth and found more paradoxes among them but also some axiomatic theories that go beyond the Friedman–Sheard criteria. Several of these were non-starters however, being omega-inconsistent or self-refuting. Each of these alethic properties, FSBtruth, FSCtruth, FSFtruth, FSItruth, FStruth, KFtruth, VFtruth, ascending truth, and descending truth, is worth investigating in detail, but the next section has space for only a few broad themes.

## 10 Are any alethic properties substantive?

Finally it is time for our study of alethic properties to start paying dividends in the discussion of deflationism. Now that we have some sense of which alethic properties exist and how they differ from what we intuitively take the property of being true to be like, we can turn to the question of what these alethic properties are like. In particular, we can think about the question of whether deflationism is right about any of these alethic properties.

Deflationism about truth has been influential in analytic philosophy for decades now, but there has been renewed interest in what deflationists should say about the



property of being true. First generation deflationists tended to deny that there was any property denoted by ‘true’, but the subsequent debate made it clear that most of them meant instead that there is no *substantive* property denoted by ‘true’.<sup>47</sup> Even so, most of the attention was on what form a deflationary theory of truth should take and on what a deflationist should say about the role of truth predicates in our linguistic practice. However, the debate has now shifted to deflationary views on the property of being true, with the result being that earlier characterizations have been seen as inadequate. There is a new sense that the debate over deflationism about truth might well be adjudicated best by evaluating what deflationists say about the property of being true.

Wyatt’s (2015) paper, “The Many (Yet Few) Faces of Deflationism,” catalogs five deflationary theses about the property of being true:

(Transparency)	Being true is a metaphysically transparent property,
(Non-explanatory)	Being true is a non-explanatory property,
(Unconstituted)	Being true is not constituted by any other property, [Wyatt claims that (Unconstituted) should not count because Horwich rejects it, but I include it because it is prominent in the literature and adds to the discussion.]
(Abundant)	Being true is an abundant property, and
(Logical)	Being true is a logical property.

These need not be accepted by every deflationist, but each has considerable support. A few clarificatory comments on these deflationary theses are in order.

First, being true is a *metaphysically transparent* property if and only if anyone who possesses the concept of truth is in a position to know all the essential facts about the property of being true. In other words, the property of being true does not have some hidden essence to be discovered by some investigation. Simply having the concept of truth is enough to be in a position to know everything important about the property of being true (but not, obviously, which things have that property).

Second, being true is a *non-explanatory* property iff there are no facts that are explained by facts about the property of being true.<sup>48</sup> It is important to stress that ‘true’ might occur in a theory only in its expressive role as a device of generalization despite the fact that the theory in question is genuinely explanatory. The word ‘true’ serves as a device of generalization when it changes a sentence position into a singular term position in a sentence. For example, one might say the following sentences. If grass is green and Sarah says that grass is green, then we should trust Sarah about this. If snow is white and Sarah says that snow is white, then we should trust Sarah about this. And so on. But one might want to state the general point and say if \_\_\_\_\_ and Sarah says that \_\_\_\_\_, then we should trust Sarah (for whatever can be put in the sentence spaces). But that isn’t a sentence. Whether

<sup>47</sup> See Beall and Armour-Garb (2005) for a survey.

<sup>48</sup> Wyatt formulates this condition in terms of facts about *the essence* of the property of being true. However, I am not clear on what this rules out. Regardless of how one formulates the condition, deflationists have long held that the property of being true can play no role in explaining meaning. As long as this result follows, nothing turns on the differences in formulation for (Non-explanatory).

English has anything that functions like those sentence spaces is contentious. But deflationists claim that the word ‘true’ helps us state the general point by turning those sentence gaps into singular term gaps, which might be filled in by a name or a description or a variable. For example: if for all things  $x$ ,  $x$  is true and Sarah says that  $x$  is true, then we should trust Sarah. The word ‘true’ turns those sentence spaces into something we can use. That is it. (Non-explanatory) entails that the word ‘true’ does not (or should not) play any role in any explanatory theory other than this generalizing role.

Third, the property of being true is *not constituted* by any other property iff it is not the case that there is some property, being  $F$ , such that all and only things that have the property of being true have the property of being  $F$ , *and* anything that has the property of being true has it *because* it has the property of being  $F$ . For example, if truth is properly analyzed as correspondence with the facts, then true things have the property of being true *because* they have the property of corresponding with the facts. Deflationists deny that truth is constituted by correspondence or by anything else. We could say that, together, (Non-explanatory) and (Unconstituted) imply that the property of being true doesn’t explain anything and nothing explains it.

A property is *abundant* iff it is to some extent unnatural, in Lewis’s sense. Naturalness, for Lewis, is objective, and it explains objective similarities among things.<sup>49</sup> Natural properties “cut nature at its joints,” in Plato’s phrase. Most metaphysicians accept that there are degrees of naturalness, but it is not clear whether deflationists endorsing (Abundant) mean that being true is merely not perfectly natural or they mean that it is highly unnatural.<sup>50</sup>

A property is *logical* iff it is invariant under certain one–one transformations of the world onto itself.<sup>51</sup> There are other ways to define ‘logical’, but Wyatt follows Tarski, who was inspired by Klein, and this invariance tradition is one of the most respected when it comes to defining what is logical. A one–one transformation of the world maps every thing in the world to some thing in the world. For example, the identity transformation maps everything to itself. If the distribution of a property is unaffected by any such transformation, then the property is logical.

What are we to make of these deflationary theses? And do any of the alethic properties canvassed count as deflationary in any of these ways?

Given what we have seen so far, it is hard to be more wrong than (Transparency). Even if we assume that anyone who possesses the concept of truth has some access to the logical platitudes, that person would probably have no idea that they were inconsistent. As we have seen, the only reasonable response to the vast amount of inconsistency among the truth platitudes is that there is no property of being true. Nothing even comes close to satisfying even small subsets of the platitudes for the concept of truth. Hence, simply possessing the concept of truth does not put one in a position to know everything essential about the property of being true.

<sup>49</sup> See Lewis (1986); see also Sider (2011).

<sup>50</sup> See Edwards (2013) and Asay (2013).

<sup>51</sup> This is very rough, see MacFarlane (2000) for a detailed treatment.

If a theorist denied this result and picked one of the alethic properties to be the property of being true, then that alethic property would be nothing like what the concept of truth would lead us to think it is like. Moreover, it has taken the accumulated effort of dozens of experts building on each others' work to figure this out. Even on a liberal reading of 'in a position to know', (Transparency) fails. For every inconsistent concept of X, if it has an associated property of being X, then being X is not a transparent property.

What about the alethic properties themselves?<sup>52</sup> For example, the property of being FSAtrue. Is it metaphysically transparent? Is anyone who possesses the concept of FSAtruth in a position to know all the essential facts about the property of being FSAtrue? This depends on what one takes to be *essential* to the property of being FSAtrue. If we assume that possessing the concept requires some kind of acceptance of the axioms in subset A of Friedman–Sheard criteria, then the answer is clearly No.<sup>53</sup> One might understand and accept these axioms without being able to figure out that there is only one model for this subset whose domain is the natural numbers and whose arithmetic vocabulary have their standard interpretations. And such a person might also not be in a position to figure out that everything is in the extension of 'true' in this model. The fact that every sentence of the object language has the property of being FSA true seems essential to me, but people differ on what is essential even in obvious cases, much less on esoteric subjects like this one. This result is even *necessary* in some sense because it holds in all the relevant models. The same sort of thing can be said for the other alethic concepts and properties. I doubt that any of them is transparent.

(Non-explanatory) should be treated as highly dubious. The concept of truth shows up in truth conditional semantic theories, which are among the most widely accepted theories for doing natural language semantics in the science of linguistics.<sup>54</sup> Denying that these theories have explanatory power would be like denying that Newton's theory of mechanics or Maxwell's theory of electrodynamics has explanatory power. It is an open question as to whether any of the alethic properties (or their related concepts) have this explanatory power, but they might work in the same sorts of semantic theories just as well as we thought truth would.<sup>55</sup>

Although we now know that there is no property of being true, the verdict on the explanatory power of the genuine alethic properties is complex. Truth conditional semantic theories come in a dizzying variety, and they are often tailored to the specific linguistic expressions under consideration. Are any of the alethic properties we canvassed up to the task? In the vast majority of cases, the clauses of a truth-conditional semantic theory could be added consistently to any of the axiomatic theories we have studied. The reason is that these semantic theories

<sup>52</sup> Thank you to a referee who emphasized this question.

<sup>53</sup> And I would say that the antecedent of this conditional is far to demanding for concept possession.

<sup>54</sup> See Chierchia and McConnell-Ginet (1990).

<sup>55</sup> Many deflationists have defended the immodest view that deflationism *is* incompatible with truth conditional semantics, but there are attempts to reconcile deflationism with a reasonable modesty toward the sciences. For example, Michael Williams' argues that the concept of truth plays only a generalizing role in truth-conditional semantic theories; see Williams (1999). See also Burgess (2011b) and McGee (2016).

are not designed to apply to language fragments that contain truth predicates. Instead, semantic theories for epistemic modals, for example, apply to language fragments that contain epistemic modals, and semantic theories for conditionals apply to language fragments that contain conditionals. The problems in which we are interested crop up only for a truth-conditional semantic theory when they are interpreted as applying to sentences in which truth predicates occur. Hence, in this sense, any of our alethic properties would be up to the task of satisfying the principles of most truth-conditional semantic theories.

But there is a catch. If we ask ourselves whether any of the alethic properties studied so far could satisfy all the principles of a truth-conditional semantic theory when it is interpreted as applying to a language with a truth predicate and the resources to construct liar sentences, then the answer is No. There are plenty of paradoxes hiding among these principles and so no alethic property is going to satisfy all of them. [For a detailed argument, see Scharp (2013)]. Therefore, any of the alethic properties could serve an explanatory role in virtually any truth-conditional semantic theory, but when it comes to a truth-conditional semantic theory for a truth predicate, none of them are up to the task. Never fear, because there is a way to fix this problem, which is a topic of the next section.

Are any of the alethic properties unconstituted? This is hard to say because it is not obvious how to individuate properties and it isn't clear what kind of dependence is invoked with the claim that the property of being true is *constituted* by some other property. If something like a reductive explanation is given as the reading of 'constituted', then the answer is probably going to be No because uncontroversial reductive explanations (think thermodynamics and statistical mechanics) are rare in philosophy. If the standard is somewhat relaxed, say, to a weak supervenience claim, then perhaps the odds are a bit better, but probably not by much. What kind of explanatorily powerful property is going to explain any one of these alethic properties? They are each hopelessly gerrymandered—their extensions zigzagging around to avoid the plethora of impossibility results. Hence it looks like all of the alethic properties are unconstituted.

One might protest this conclusion: but we already have such properties! Correspondence to the facts or coherence, or superwarrant maybe. My reply: none of the classic analyses of truth are remotely plausible in light of the results of our attempted Canberra plan. First of all, there is no property of being true. That result is inconsistent with every purported constitution theory for the property of being true. Second, even if one denies this conclusion, it is hard to see that the property of corresponding to the facts could constitute any one of the alethic properties. Consider the property of being FSFtrue (i.e., satisfying subset F of Friedman–Sheard criteria). This property fails to obey ( $\neg$ T-Enter). Hence, even if one can prove a theorem from subset F of form  $\neg p$ , one cannot conclude that  $\neg Tp$ . For example, it would be like proving 'snow is not white' but failing to prove 'snow is white' is not true'. Does the property of corresponding to the facts behave just like this? Highly doubtful.

What about abundance? All of the alethic properties are highly unnatural in one sense. Naturalness is supposed to explain the objective similarities between things. If we intuitively think of all the truths as having some kind of objective similarity,

then every alethic property is going to violate this in myriad ways. And it is hard to believe that anyone thinks that all the KFtruths or all the VFtruths have objective similarity. So in this sense all the alethic properties are deflationary.

Logicity? Wyatt argues that no deflationist should accept this thesis because it is easily refuted.<sup>56</sup> His argument is that some transformations will map some true proposition onto a false proposition, so truth is not preserved under all transformations. But Wyatt assumes that propositions are things in a particular world, whereas it is much more common to assume that propositions are sets of possible worlds, and so they are not members of any particular possible world. If that is right, then propositions are not among the things in a world, and so Wyatt's argument fails.

Certainly some alethic properties count as logical: when the predicates that denote them get their own clauses in the semantic theory for languages in which they occur, just like negation and the rest of the logical vocabulary. If that is the case, then the interpretation of the predicate is by definition invariant across all models. For example, one can do this for ascending truth and descending truth.<sup>57</sup> So at least some alethic properties count as logical properties.

We can summarize the results in this section:

- i. No alethic properties are transparent, so versions of deflationism committed to (Transparency) are false.
- ii. No alethic properties are non-explanatory, so versions of deflationism committed to (Non-Explanatory) are false.
- iii. All alethic properties are unconstituted (depending on how one individuates properties), so insofar as they are committed to (Unconstituted), deflationist views of alethic properties are acceptable.
- iv. All alethic properties are abundant, so insofar as they are committed to (Abundant), deflationist views of alethic properties are acceptable.
- v. Some alethic properties are logical (depending on how one understands logical expressions and how one formulates a semantic theory for a language with expressions that denote these properties), so insofar as they are committed to (Logical), deflationist views of alethic properties might be acceptable.

It deserves to be emphasized that deflationist theories of truth have multiple aspects or parts, some of which pertain to the property of being true, some to the concept of truth, some to the word 'true', and some to the structure of any acceptable theory of truth. Moreover, there is no property of being true, so any theory, deflationist or not, that entails that there is such a property is false. At present, we are evaluating only deflationist views of alethic properties, not deflationist views as a whole. As such, even those versions of deflationism that come out as acceptable on the present inquiry might be false for some other reason (e.g., because deflationist theories are typically taken to consist of all and only the non-paradoxical instances

<sup>56</sup> Wyatt (2015: pp. 16–17).

<sup>57</sup> See Scharp (2013).

of (Schema T), but it is difficult or impossible to specify in advance which instances these will be).

## 11 New alethic concepts

There is a meaningful English word, ‘true’, and there is a concept of truth expressed by that word, but the concept of truth is defective in the sense that its platitudes are inconsistent. Moreover, they are seriously inconsistent—the average size of the nine maximally consistent subsets of twelve Friedman–Sheard criteria is only six, and the average size of the somewhat reasonable ones is only five. As such, there is no property of being true. Still, there are plenty of properties that are somewhat like what we think the property of being true should be like. These are the alethic properties, and there is one for each consistent (and omega-consistent) subset of platitudes for the concept of truth.

Instead of ending our inquiry here, we have decided to see whether we can improve our situation by adding new alethic *concepts* to our conceptual scheme. Our guide to the world of new alethic concepts is the realm of alethic properties. In other words, for each alethic property, there is a new alethic concept, distinct from the concept of truth. There are, of course, lots of other concepts like truth but distinct from it, but the ones based on alethic properties come with a guarantee: *I am not inconsistent*.<sup>58</sup> Each new alethic concept has, as its constitutive principles, all the platitudes in the theory it satisfies—e.g., the concept of KFtruth has as its platitudes all the axioms of the theory KF, which is satisfied by the property of being KFtrue.

Which of these new alethic concepts should we use in place of truth? There is good reason to think that no single new alethic concept is up to the task because no single alethic concept can do everything we expect the concept of truth to do. For example, it is widely accepted that we use ‘true’ to endorse propositions that we cannot assert directly; for example, Ralph can assert ‘the Riemann hypothesis is true’ and thereby endorse the Riemann hypothesis even though he does not remember or never learned which sentence expresses it. Or he can assert ‘all the axioms of ZFC are true’ and thereby endorse all the axioms of ZFC even though there are too many for him to assert one by one. We can capture this role by saying that a truth predicate functions as *device of endorsement*. The flip side of this role is a *device of rejection*; he can say ‘the Riemann hypothesis is not true’ and thereby reject the Riemann hypothesis. In order to serve as a device of endorsement, the truth predicate must obey (T-Out), and in order to serve as a device of rejection, the truth predicate must obey (T-In). Of course, we already know that in a classical setting no single concept obeys these two principles; thus, no concept can serve as both a device of endorsement and rejection given classical logic and the expressive resources to construct liar sentences. However, if we replace truth with two concepts, we can split

<sup>58</sup> And an even further guarantee against consistent but omega-inconsistent concepts. As far as I know, no one has considered omega-inconsistent concepts before.

the workload, allowing one to serve as a device of endorsement and the other to serve as a device of rejection.

Moreover, anyone who advocates replacing truth with a single alethic concept would have to pick the best combination of truth platitudes and give up anything like the rest of them. However, when we replace truth with *two* concepts, we have the option of accepting some platitudes that are formulated with one concept and some platitudes that are formulated with the other. In addition, this strategy allows for *hybrid* platitudes, which are formulated with both concepts (e.g., a hybrid platitude would be ‘If a conditional is *descending true*, then if the antecedent is *descending true* then the consequent is *ascending true*’). The theory of the replacement concepts I offer includes a replacement principle for *every one* of the platitudes listed above for truth. Such a thing is only possible when we replace truth with multiple new alethic concepts.

One significant issue for choosing between new alethic concepts is whether the alethic properties they denote are deflationary or substantive, and in what sense. We saw five ways of drawing the deflationary/substantive distinction for the alethic properties from Wyatt, and conducted a preliminary investigation into which alethic properties have these features. Against the background of a conceptual engineering project, the question becomes: what do we want our replacement concepts to be like? What are the considerations for and against various features—explanatory power, expressive power, naturalness? And which ones do we want, given what we are interested in doing with our replacement concepts?

Another problem surfaced in the previous section: no new alethic concept, by itself, can serve the explanatory role we save for the concept of truth. To make the point, focus on truth-conditional semantics. Recall that any of the alethic properties could serve the explanatory role required by the vast majority of semantic theories in the truth-conditional tradition because few of these theories are intended to apply to fragments of natural language that include a truth predicate. However, none of them can be used to formulate a successful truth-conditional semantic theory that applies to the fragment of a natural language containing a truth predicate. This should be obvious: the alethic properties are nothing like what we think of as the property of being true, so the alethic properties deliver nothing like what we think of as truth conditions.

However, if we adopt ascending truth and descending truth into our conceptual scheme, then we can formulate a successful semantic theory using them. The theory specifies *ascending* truth conditions and *descending* truth conditions for all the sentences in a fragment of natural language that contains a truth predicate, an ascending truth predicate, and a descending truth predicate. Moreover, for safe sentences—roughly, the non-paradoxical ones—the ascending truth conditions are identical to the descending truth conditions, which are identical to the truth conditions. Thus, *ascending and descending truth conditional semantics* reduces to *truth-conditional semantics* in all the familiar cases where the distinction between ascending truth and descending truth is negligible. That is similar to the situation in physics—relativistic mechanics reduces to Newtonian mechanics in all the familiar cases where the distinction between relativistic mass and proper mass is negligible. For details, see Scharp (2013).



Whether my suggested replacements for the concept of truth can actually do all the things we think truth should be able to do is still an open question. Truth's expressive role is complex and truth's explanatory role is vast. Here I have focused only on the expressive roles of acceptance rejection, and generalization, and the explanatory role in contemporary natural language semantics. As such, we have just scratched the surface of the conceptual engineering project for truth.

What we desperately need is a catalogue of all the inconsistencies among all the logical platitudes for truth listed in section two and a catalogue of all the maximal consistent subsets of these criteria. John Burgess has already suggested doing something like this, so we can reiterate his call to arms.<sup>59</sup> We can call it the *aletheic platitudes project* (APP). When complete, APP would be a gold mine for the study of aletheic properties. Moreover, APP would finally allow us to complete the conceptual engineering project of finding the best team of new aletheic concepts to replace our defective concept of truth. I think the community of mathematically minded philosophers and logicians can work together to make APP a real success. We can do this, and you can help!

**Acknowledgements** Thank you to Patrick Greenough, Michael Lynch, Nathan Kellen, Alison Duncan Kerr, Jeremy Wyatt, and several anonymous referees for their helpful comments on earlier drafts.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Asay, J. (2013). *The primitivist theory of truth*. Cambridge: Cambridge University Press.
- Bar-On, D., & Simmons, K. (2007). The use of force against deflationism. In D. Greimann & G. Siegart (Eds.), *Truth and speech acts*. New York: Routledge.
- Beall, J. (2009). *Spandrels of truth*. Oxford: Oxford University Press.
- Beall, J. C., & Armour-Garb, B. (Eds.). (2005). *Deflationism and paradox*. Oxford: Oxford University Press.
- Beall, J., & Glanzberg, M. (2008). Where the paths meet: Remarks on truth and paradox. *Midwest Studies in Philosophy*, 32, 169–198.
- Belnap, N., & Grover, D. (1973). Quantifying in and out of quotes. *Studies in Logic and the Foundations of Mathematics*, 68, 17–47.
- Braddon-Mitchell, D., & Robert, R. (Eds.). (2009). *Conceptual analysis and philosophical naturalism*. Cambridge, MA: MIT Press.
- Burgess, A. (2006). *Identifying fact and fiction*. Ph.D. Dissertation, Princeton University.
- Burgess, A. (2011b). Mainstream semantics+deflationary truth. *Linguistics and Philosophy*, 34, 397–410.

<sup>59</sup> Burgess (2011a).



- Burgess, J. (2011a). Friedman and the axiomatization of Kripke's theory of truth. In N. Tennant (Ed.), *Foundational adventures: Essays in honor of Harvey M. Friedman*. London: Templeton Press.
- Burgess, A., & Plunkett, D. (2013a). Conceptual ethics I. *Philosophy Compass*, 8, 1091–1101.
- Burgess, A., & Plunkett, D. (2013b). Conceptual ethics II. *Philosophy Compass*, 8, 1102–1110.
- Cappelen, H. (2018). *Fixing language: Conceptual engineering and the limits of revision*. Oxford: Oxford University Press.
- Chierchia, G., & McConnell-Ginet, S. (1990). *Meaning and grammar: An introduction to semantics*. Cambridge: MIT Press.
- Cieśliński, C. (2017). *The epistemic lightness of truth*. Cambridge: Cambridge University Press.
- Cook, R. (2014). *The Yablo paradox*. Oxford: Oxford University Press.
- Devlin, J. (2003). An argument for the error theory of truth. *Philosophical Perspectives*, 17, 51–82.
- Edwards, D. (2013). Truth as a substantive property. *Australasian Journal of Philosophy*, 91, 271–294.
- Eklund, M. (2002). Inconsistent languages. *Philosophy and Phenomenological Research*, 64, 251–275.
- Field, H. (2004). The consistency of the naïve theory of properties. *The Philosophical Quarterly*, 54, 78–104.
- Field, H. (2008). *Saving truth from paradox*. Oxford: Oxford University Press.
- Field, H., Lederman, H., & Fjetland, T. (2017). Prospects for a naïve theory of classes. *Notre Dame Journal of Formal Logic*, 58, 461–506.
- Fischer, M., Halbach, V., Kriener, J., & Stern, J. (2015). Axiomatizing semantic theories of truth? *Review of Symbolic Logic*, 8, 257–278.
- Friedman, H., & Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33, 1–21.
- Friedman, H., & Sheard, M. (1988). The disjunction and existence properties for axiomatic systems of truth. *Annals of Pure and Applied Logic*, 40, 1–10.
- Georgi, G. (2015). A propositional semantics for substitutional quantification. *Philosophical Studies*, 172, 1183–1200.
- Greenough, P. (2001). Free assumptions and the liar paradox. *American Philosophical Quarterly*, 38, 115–135.
- Halbach, V. (2011). *Axiomatic theories of truth*. Cambridge: Cambridge University Press.
- Halbach, V., & Horsten, L. (2015). Norms for theories of reflexive truth. In H. Galinon, T. Achourioti, & F. Fujimoto (Eds.), *Unifying the philosophy of truth*. Berlin: Springer.
- Hand, M. (2007). Objectual and substitutional interpretations of the quantifiers. In D. Jacquette (Ed.), *Philosophy of logic*. Amsterdam: Elsevier.
- Haslanger, S. (2000). Race and gender: (What) are they? (what) do we want them to be? *Nous*, 34, 31–55.
- Horsten, L. (2011). *The Tarskian turn*. Cambridge, MA: MIT Press.
- Horwich, P. (1998). *Truth* (2nd ed.). Oxford: Clarendon Press.
- Hugly, P., & Sayward, C. (2002). There is a problem with substitutional quantification. *Theoria*, 68, 4–12.
- Jackson, F. (1997). *From metaphysics to ethics: A defense of conceptual analysis*. Oxford: Oxford University Press.
- Johnston, M., & Leslie, S.-J. (2012). Concepts, analysis, generics and the Canberra plan. *Philosophical Perspectives*, 26, 113–171.
- Ketland, J. (2005). Yablo's paradox and  $\omega$ -inconsistency. *Synthese*, 145, 295–302.
- Kripke, S. (1976). Is there a problem about substitutional quantification? In G. Evans & J. McDowell (Eds.), *Truth and meaning*. Oxford: Oxford University Press.
- Künne, W. (2003). *Conceptions of truth*. Oxford: Clarendon Press.
- Leigh, G. (2015). Some weak theories of truth. In K. Fujimoto, H. Galinon, & T. Achourioti (Eds.), *Unifying the philosophy of truth*. Berlin: Springer.
- Leigh, G., & Rathjen, M. (2010). An ordinal analysis for theories of self-referential truth. *Archive for Mathematical Logic*, 49, 213–247.
- Leitgeb, H. (2007). What theories of truth should be like (but cannot be). *Philosophical Compass*, 2, 276–290.
- Lewis, D. (1970). How to define theoretical terms. *Journal of Philosophy*, 67, 427–446.
- Lewis, D. (1973). *Counterfactuals*. New York: Wiley.
- Lewis, D. (1986). New work for a theory of universals. *Australasian Journal of Philosophy*, 61, 343–377.
- Lynch, M. (2009). *Truth as one and many*. Oxford: Oxford University Press.
- MacFarlane, J. (2000). *What Does it mean to say that logic is formal?* Ph. D. dissertation, University of Pittsburgh.
- Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York: Penguin.

- McGee, V. (1985). How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14, 399–410.
- McGee, V. (1991). *Truth, vagueness and paradox*. Hackett.
- McGee, V. (2016). Thought, thoughts, and deflationism. *Philosophical Studies*, 173, 3153–3168.
- Montague, R. (1963). Syntactical treatments of modality, with corollaries on reflection principles and finite axiomatizability. *Acta Philosophica Fennica*, 16, 153–167.
- Novaes, C. D. (2018). Carnapian explication and ameliorative analysis: A systematic comparison. *Synthese*. <https://doi.org/10.1007/s11229-018-1732-9>.
- Novaes, C. D., & Reck, E. (2017). Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization. *Synthese*, 194, 195–215.
- Olson, J. (2014). *Moral error theory: History, critique, defense*. Oxford: Oxford University Press.
- Priest, G. (2006). *In contradiction* (2nd ed.). Oxford: Oxford University Press.
- Scharp, K. (2013). *Replacing truth*. Oxford: Oxford University Press.
- Scharp, K. (2018). Replies to Bacon, Eklund, and Greenough on replacing truth. *Inquiry*. <https://doi.org/10.1080/0020174X.2017.1420012>.
- Scharp, K. (2020). Philosophy as the study of defective concepts. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics*. Oxford: Oxford University Press.
- Schindler, T. (2015). A disquotational theory of truth as strong as  $Z_2^-$ . *Journal of Philosophical Logic*, 44, 395–410.
- Schroeder, M. (2010). How to be an expressivist about truth. In C. Wright & N. Pedersen (Eds.), *New waves in truth*. New York: Palgrave.
- Sheard, M. (1994). A guide to truth predicates in the modern era. *The Journal of Symbolic Logic*, 59, 1032–1054.
- Sheard, M. (2002). Truth, probability, and naive criteria. In V. Halbach & L. Horsten (Eds.), *Principles of truth*. Frankfurt a. M.: Dr. Hänsel-Hohenhausen.
- Sider, T. (2011). *Writing the book of the world*. Oxford: Oxford University Press.
- Tarski, A. (1933). The concept of truth in formalized languages. In T. Woodger & J. Corcoran (Eds.), *Logic, semantics, meta-mathematics*. Indianapolis: Hackett Publishing Company.
- Vann, M. (1992). Maximal consistent sets of instances of Tarski's schema (T). *Journal of Philosophical Logic*, 21, 235–241.
- Williams, M. (1999). Meaning and deflationary truth. *Journal of Philosophy*, 96, 545–564.
- Wyatt, J. (2015). The many (yet few) faces of deflationism. *Philosophical Quarterly*, 263, 362–382.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.